



Peter Kabel

Dialog zwischen Mensch und Maschine

Conversational User Interfaces,
intelligente Assistenten und
Voice-Systeme



Springer Gabler

Dialog zwischen Mensch und Maschine

Peter Kabel

Dialog zwischen Mensch und Maschine

Conversational User Interfaces, intelligente
Assistenten und Voice-Systeme



Springer Gabler

Peter Kabel
Hamburg, Deutschland

ISBN 978-3-658-29584-4 ISBN 978-3-658-29585-1 (eBook)
<https://doi.org/10.1007/978-3-658-29585-1>

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer Gabler
© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2020
Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.
Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.
Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag, noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Springer Gabler ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Vorwort

Es wird nicht mehr lange dauern und wir werden uns wundern, wenn ein Gerät in unserem Haushalt oder im Büro nicht auf unsere Sprache reagiert. Mit Möbeln, Lautsprechern, Fernsehern oder Waschmaschinen zu reden, wird so selbstverständlich sein, wie mit einem Mitmenschen zu sprechen. Sprache ist für uns Menschen die natürlichste und am intensivsten genutzte Form der Kommunikation.

Im Dialog mit Maschinen waren grafische Benutzeroberflächen, ob nun per Maus oder per Finger bedient, eigentlich ein technischer Umweg – allerdings ein notwendiger. Sprache war lange Zeit zu komplex, als dass Maschinen sie übersetzen und verstehen konnten. Das hat sich geändert und darum stehen wir vor einer Disruption, deren Bedeutung durchaus vergleichbar mit der Einführung des Internets oder auch des Smartphones ist. Conversational User Interfaces (CUIs) sind das nächste große Ding.

Die entscheidende Neuerung: Die Systeme haben gelernt, den Menschen zu verstehen. Wir müssen nicht mehr bestimmte Formulierungen nutzen oder Befehle in festgelegter Reihenfolge aufsagen. Wir reden mit der Maschine, wie wir mit unserem menschlichen Gegenüber sprechen. Natürliche Sprache ist das Stichwort. Das umfasst gesprochene Sätze, aber auch geschriebene Sätze, beispielsweise auf einer Smartphone- oder Computertastatur. In Anwendungen, die wie ein Messenger funktionieren, führen wir einen Dialog mit einem Computer. So buchen wir eine Fahrkarte, verlängern einen Handyvertrag oder starten eine Überweisung. Innerhalb kurzer Zeit werden die Maschinen in der Lage sein, immer mehr Aufgaben zu erledigen. Die Entwickler werden immer mehr Schnittstellen zu Informations-, Service- und Verkaufsangeboten öffnen.

Das bedeutet nicht nur eine Umgewöhnung für die Nutzer, sondern auch für die Anbieter: Wie präsentiert man Produkte, die man nicht sieht? Die Herausforderungen für Online-Händler sind vielfältig: Das beginnt beim Aufbau einer Produktdatenbank nach semantischen Suchkriterien und reicht bis zum Audio-SEO. Die Herausforderungen sind gewaltig, denn wie beim Aufkommen des Internets und später von Mobile, wird sich erneut das gesamte Ökosystem ändern. Auf Verbraucherseite stellt sich die Frage: Werden Kunden Produkte bestellen, die sie nie zu Gesicht bekommen haben? Das eröffnet das weite Themenfeld „Vertrauen“ – quasi blindes Vertrauen. Es wird interessant zu sehen, wie die Anbieter damit umgehen oder besser gesagt, wie sie es umgehen. Gleichzeitig dürfen die

Anbieter das Thema Sicherheit nicht unterschätzen. In diesem Buch lernen Sie etwas über Adversarial Attacks. Ein erschreckend beeindruckende Methode, um CUIs in die Irre zu führen. Wie bei fast jedem technischen System entsteht hier ein Katz-und-Maus-Spiel zwischen Anbietern und Betrügern.

Gleichzeitig werden die Maschinen immer schneller lernen und immer besser erkennen, ob sie in die Irre geleitet werden sollen. Sie werden Muster und Routinen erkennen und proaktiv Dienste für ihre Besitzer aktivieren. Das Erkennen von Stimmungen und Gefühlszuständen ist kein allzu weit entferntes Ziel. Maschinen werden dann entsprechend darauf reagieren.

Viele mögen im ersten Moment von einer derartigen Zukunft erschreckt sein. Doch sie wird kommen. Allein wenn man sich die Entwicklungen in Sachen Finanzen und Aufbau der Entwicklerteams bei den beiden großen Anbietern, Google und Amazon, anschaut, wird deutlich: Die Zukunft gehört der Sprache. Sie wird die dominierende Schnittstelle zu jeglicher Art von Maschine.

Noch stehen die Ökosysteme ganz am Anfang. Vieles mag noch nicht richtig funktionieren oder durch die Marketingabteilung etwas zu positiv beschrieben werden. Doch das Tempo ist atemberaubend. Sie und Ihre Organisation sollten hier den Anschluss nicht verpassen und besser von Anfang an dabei sein. Denn Conversational User Interfaces sind das nächste große Ding.

Um die Inhalte dieses Werkes noch besser zu durchdringen und zu verstehen gibt es zu diesem Buch auch einen Fragen+Antworten-Datensatz in der Springer-Nature-Flashcards-App. Hierzu können Sie einfach den Code aus diesem Buch in die kostenlose App eingeben und anschließend auf Ihrem Smartphone mit Fragen und Antworten zum Buch das Gelesene überprüfen und verinnerlichen – innovativ und interaktiv, wie Conversational User Interfaces selbst. Ein Tipp: Das Glossar am Ende des Buches wird Ihnen dabei helfen.

Wir sprechen uns.

Hamburg, Deutschland
April 2020

Peter Kabel

Danksagung

Ein besonderer Dank geht an die Beteiligten des Forschungsprojekts www.tell-me.io:
Lisa Beranek, Marlena Reinke, Tom-Lucas Säger, Partrick Surdziel, Felix Zepke sowie
an Ann-Kristin Iwersen und Dirk Kunde.

Inhaltsverzeichnis

1 Was sind Conversational User Interfaces?	1
1.1 Der Wunsch, mit Maschinen zu sprechen	3
1.2 Instant Messenger als Vorläufer heutiger Chatbots	5
1.3 Das Voice Jail der Telefon-Hotlines	6
1.4 Darum eignet sich Sprache für einen Dialog mit Maschinen	7
1.5 Trainingsdaten in digitalisierter Form	8
1.6 Das Versprechen von CUIs und intelligenten Assistenten	10
1.7 Die Qual der Wahl – Finden statt suchen	11
1.8 Herausforderungen von CUIs und intelligenten Assistenten	12
1.9 Multimodal könnte in vielen Anwendungsfällen die Lösung sein	15
Literatur	15
2 Stand heute: Sechs Beispiele für funktionierende CUIs	17
2.1 Google Assistant und Amazon Alexa	18
2.2 LH Group Chatbot: Mein Flug wurde gestrichen	20
2.3 Maybe: Lily bringt dem Zuhörer Chinesisch bei	21
2.4 Niki.ai: Eine Busfahrkarte kaufen	23
2.5 Nomi: Wo soll die Fahrt hingehen?	24
2.6 Param: Den passenden Bewerber finden	26
Literatur	26
3 Woraus setzen sich CUIs und intelligente Assistenten zusammen?	29
3.1 Wo findet nun Artificial Intelligence statt?	32
3.2 Was muss ich bei Planung und Aufbau eines CUI oder Assistenten beachten?	32
3.2.1 Branding, Naming und Gestalt	32
3.2.2 Wie intelligent soll Ihr Assistent werden?	35
3.2.3 Unterhaltung und Funktionalität	36
3.2.4 Interaktion und Dialog	36
3.2.5 Monetarisierung und Bezahlung	38
3.2.6 Monitoring und Reporting	38
Literatur	38

4 Was sind NLP und NLG und wie funktionieren sie?	39
4.1 NLP-Anwendungen nach Branchen	44
4.2 Wie funktioniert Natural Language Understanding?	46
4.3 Was ist Machine Learning?	46
4.4 Gängige NLP-Techniken	48
4.5 Wie funktioniert Voice to Text?	52
4.6 Erzeugung von Texten – Natural Language Generation	57
4.7 Aktuelle Herausforderungen in NLP	59
4.8 NLP für Nicht-Englisch	60
4.9 Ausblick	61
Literatur	62
5 Weshalb sind CUIs und intelligente Assistenten so bedeutend?	65
5.1 Warum jetzt? Und weshalb so schnell?	66
5.2 Der Wettlauf hat längst begonnen	68
5.3 Google knüpft ein weitgespanntes Netz	69
5.4 Amazon: Marktführer bei Smartspeakern	71
5.5 Apple: Siris Frühstart	73
5.6 Offline ist das neue Online	74
5.7 Conversational Productivity Tools – Apple und Microsoft konkurrieren wie in alten Desktop-Zeiten	74
5.8 Marketingabteilungen als Temperaturfühler	78
5.9 Manches kann, vieles muss sich ändern	78
5.10 Neue Regeln im Zeitalter der Absichten	79
5.11 Die Veränderung ist massiver, als man zunächst denkt	80
Literatur	81
6 Das sind die Player	83
6.1 Die Plattformen, CUI-Kanäle und -Endgeräte	84
6.2 CUI-Plattformen – Software-Integratoren und die Allrounder	86
6.3 CUI-Backends und Businesslösungen	87
6.4 Open Source und On-Premisse	89
6.5 Die CUI-App-Anbieter	89
6.6 Die Dienstleister	90
6.7 Testen, Optimieren und Prototypen	91
6.8 Lösungen in Kombination aus Hard- und Software	92
Literatur	95
7 Look East	97
7.1 Demografie und Sprachen	98
7.2 Bocksprünge	99
7.3 Boomende Annahme	100
7.4 Indische Sprachvielfalt	101

7.5	Mehrdeutigkeit und Komplexität der Sprachen	101
7.6	Mangel an sprachlicher Grammatik, Literatur und dokumentierten Standards	102
	Literatur	103
8	Tech Stack: Wie die erfolgreiche Umsetzung im Unternehmenskontext gelingt – und warum das gesamte Unternehmen betroffen ist	105
8.1	Technologielandschaft	106
8.2	CUI-Clients und -Endgeräte	107
8.3	CUI-Channel (Messaging- & Digital-Assistant-Dienste)	107
8.4	CUI-Plattformen	109
8.5	CUI-Backend	110
8.6	Technische Herausforderungen	111
8.6.1	Reichweite	111
8.6.2	Verständlichkeit	112
8.6.3	Intelligenz	113
8.6.4	Themen-Bandbreite	113
8.6.5	Relevanz	114
8.7	Projektumsetzung	114
8.8	CUI-Pilot	116
8.9	CUI-Transformation	116
8.10	Wann ist ein CUI erfolgreich?	117
8.11	Fazit	118
	Literatur	118
9	So verändern sich Ökosysteme	119
9.1	Die Ausdehnung des Universums	122
9.2	Bundling und Un-Bundling	122
9.3	Von der App zur Micro-App	123
9.4	Voice Search	126
9.5	Multiple Plattform	126
9.6	Welche Bedeutung haben Websites künftig noch?	128
9.7	Hinter jeder Suche steht ein Bedürfnis	129
9.8	Explizit und implizit	130
9.9	Auffinden von Services und Produkten	132
9.10	Findability und Discoverability	134
9.11	Customer Journey Commerce	135
9.12	Berührungsloser Konsum	136
9.13	Mentale Adipositas	137
9.14	Walled Gardens vs. offene Architekturen	137
9.15	Transaktionen	139
	Literatur	140

10 Reality Check	143
10.1 Erwartungsmanagement	144
10.2 Herausforderung Multimodalität	145
10.3 Woher stammt das notwendige Wissen?	145
10.4 Stockende Dialoge	147
10.5 Fehlende Emotionen	149
10.6 Mangelnde Mehrsprachigkeit	150
10.7 Grobe Werkzeuge	150
10.8 Wo stehen wir auf dem Hype Cycle?	150
10.9 Voreingenommenheit und Fehler von Algorithmen	151
10.10 Datenschutz und Datensicherheit	152
10.11 Adversarial Attacks	153
10.12 Schnelle Schritte	155
Literatur	156
11 Branchen und Use Cases	159
11.1 Sprache als universelles Interface	159
11.2 Eignet sich jeder Anwendungsfall für Sprache?	160
11.3 Eignet sich jeder Anwendungsfall für einen intelligenten Assistenten? ..	162
11.4 Strukturelle Sweetspots	163
11.5 Software-Agenten, die erinnern und aufmerksam machen	163
11.6 Intelligente Assistenten, die etwas für Menschen erledigen	164
11.7 Bots, die Inhalte zusammentragen und vorlegen	165
11.8 Bots, die Menschen prüfen und Sachverhalte erfassen	166
11.9 Assistenten, die zwischen Menschen vermitteln	170
11.10 Bots, die aktiv in Dialog gehen	171
11.11 Aus Wearables werden Awareables	172
Literatur	173
12 Machen: Der Friesenbot	175
Glossar	189



Was sind Conversational User Interfaces?

1

Zusammenfassung

Nicht alle Conversational User Interfaces (CUI) sind gleich. Der Leistungsumfang variiert stark, so dass die Begriffsdefinition in diesem Kapitel zunächst für Klarheit sorgt. Ein Blick in die Geschichte zeigt, dass der Mensch schon lange mit Maschinen bzw. menschenähnlichen Abbildern sprechen möchte. Sprache ist die natürlichste Kommunikationsform. Doch mit Maschinen mussten wir erst Umwege über Befehlszeilen, grafische Benutzeroberflächen bis hin zur Spracherkennung bei telefonischen Hotlines machen. Mit CUIs kann man Hierarchieprobleme klassischer Informationsanordnungen umgehen. Dafür entstehen neue Herausforderungen wie das Auswahl- und Vertrauensdilemma.

Die Herausforderung beginnt bereits bei der Übersetzung: Conversational User Interfaces (CUI) sind mehr als nur Sprachassistenten. Bei letzteren denkt jeder sofort an Googles Assistant, Apples Siri, Amazons Alexa, Microsofts Cortana und vielleicht noch Samsungs Bixby. Wie bei allen Interfaces ist das CUI nur die Oberfläche, eben der Einstieg. In diesem Fall passiert der Einstieg über natürliche Sprache. Dahinter liegen unterschiedliche technische Systeme, für die ein bunter Mix an Bezeichnungen verwendet wird: Bots, Chatbots, Voicebots oder (intelligente) Assistenten.

Dabei umfasst unsere Begriffsdefinition für ein CUI sämtliche natürlichsprachige Kommunikation zwischen Mensch und Maschine. Mit natürlichsprachig ist nicht nur das gesprochene Wort gemeint. Auch geschriebene Sprache zählt dazu. Insbesondere auf mobilen Geräten wie Smartphones und Tablets verschwimmen die Grenzen. Spricht der Nutzer in sein Gerät, führt Automatic Speech Recognition zu einem geschriebenen Text. Diesen analysiert die Maschine mithilfe von Verfahren des Natural Language Processing

Entwicklungsstufe	Sprachbehandlung	Servicemöglichkeiten
Basic Bot	Parsing, einzelne Frage, einzelne Antwort	Bereitstellung von Text, Link, Bild. Keine Systemintegration
Standard Bot	NLP, Dialogmanagement	Zugriff auf lokale Daten, Zugriff auf externe Daten, einfache Transaktionen
Advanced Bot	Contextual NLP, Machine Learning, Dialogmanagement, User Präferenzen	Zugriff auf lokale und externe Daten, komplexe Transaktionen

Abb. 1.1 Klassifizierung von Bots

(NLP). Beide Formen sind für den Menschen natürliche Eingabeformate und werden von CUIs genutzt. Da NLP unterschiedliche Technologien umfasst, die auch Machine-Learning-Methoden enthalten, werden CUIs häufig auch „Conversational AI“ genannt.

In diesem Buch unterscheiden wir drei Entwicklungslevel: Basic, Standard und Advanced (siehe Abb. 1.1). Der **Basic Bot** nutzt Parsing zur Erkennung der Sprache. Dabei handelt es sich in der Regel um Schriftsprache. In den Sätzen wird nach Schlüsselwörtern gesucht, die daraufhin eine Aktion auslösen. Taucht das Wort „Kündigung“ auf, erhält der Fragende eine Antwort zu Kündigungsfristen oder es wird eine Vertragskündigung ausgelöst.

Der **Standard Bot** analysiert gesprochene Sprache bereits mit einem NLP-System. Somit kann er herausfinden, was das Anliegen (Intent) des Nutzers ist. Daraufhin nutzt der Standard Bot Schnittstellen zu internen als auch externen Systemen, um die gewünschten Antworten zusammen zu stellen. Diese wird in Schriftsprache oder per Text-to-Speech als gesprochene Sprache ausgegeben. Hier sind bereits erste Dialoge und durch die Anbindung von IT-Systemen auch Transaktionen möglich.

Der **Advanced Bot** verdient die Bezeichnung Sprachassistent. Zur Erkennung des Anliegens (Intent) erkennt dieser Bot auch Zusammenhänge und thematische Hintergründe (Context). Er weiß, dass sich die Farbe Grün auf den zuvor gewünschten Pullover bezieht. Mit dem Advanced Bot ist bereits ein richtiger Dialog möglich. Der Bot ist deutlich besser im Verstehen von Anliegen, da er über Machine-Learning-Methoden große Mengen an Trainingsdaten verarbeitet hat. Beispielsweise hat ein Versandhändler zur Abwicklung von Reklamationen die Abschriften bzw. Mitschnitte aus dem Call-Center zu Trainingszwecken eingesetzt. Ist der Nutzer über sein Konto beim Bot-Anbieter angemeldet, kann dieser auf frühere Bestellungen, Vorlieben und Gewohnheiten eingehen. Ein Advanced Bot nutzt diverse Schnittstellen zu internen und externen Serviceangeboten, um Transaktionen und Services anbieten zu können. Die Ausgabe der Antwort erfolgt auf gleichem Weg wie die Eingabe – entweder in Schriftform oder in gesprochener Sprache.

Neben den Unterschieden in den dahinter liegenden Schnittstellen zur Business Intelligence sowie allen Formen der natürlichen Sprache gibt es auch Unterschiede bei der Hardware. Die lässt sich mit dem Begriff Multimodalität am besten beschreiben. Dabei geht es

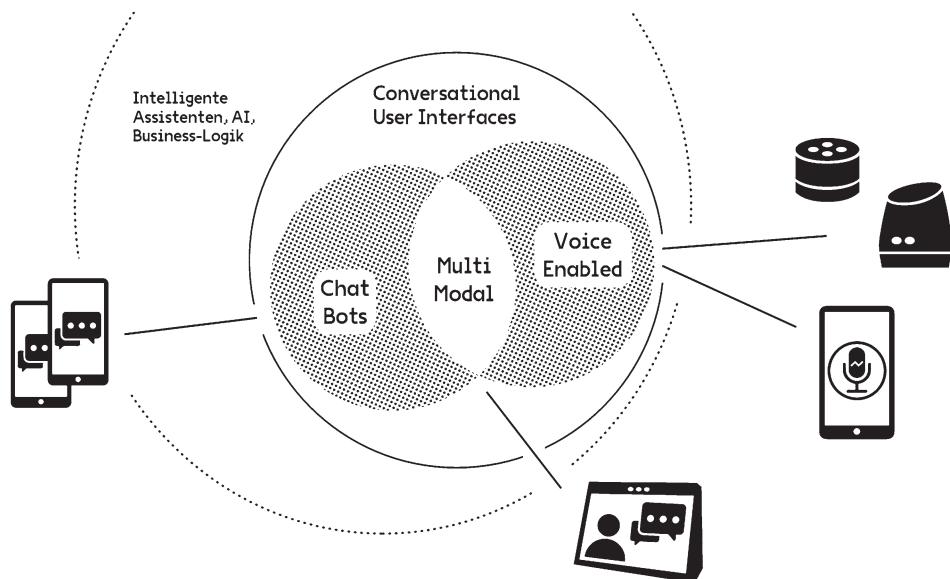


Abb. 1.2 Der Conversational-User-Interfaces-Kosmos (CUI)

um die Informationsausgabe auf unterschiedlichen Geräten. Schriftsprache, eine Auswahl an Optionen als auch Produktvorschläge sollte man in manchen Fällen nicht nur hören, sondern auch sehen. Mit der optischen Information fällt es leichter, eine Entscheidung zu treffen. Dazu benötigt man also ein Gerät mit Bildschirm. Darum nutzen viele Geräte, allen voran Smartphones, ein Display im Zusammenspiel mit einem CUI (Abb. 1.2).

1.1 Der Wunsch, mit Maschinen zu sprechen

Bereits Jahrhunderte vor dem ersten Chatbot dachten Menschen über Abbilder ihrer selbst nach. In der jüdischen Literatur findet sich ab dem frühen Mittelalter die Geschichte des Golems. Es ist ein großes menschliches Wesen, das aus Lehm geformt wurde. Der Golem sollte Arbeiten übernehmen, die der Mensch nicht erledigen wollte. In Mary Shelleys Roman von 1818 baut Victor Frankenstein in seinem Labor ein großes, menschliches Monster – hässlich, aber durchaus sensibel und voller Gefühle. Auch der körperlose HAL 900 aus Stanley Kubricks „2001: Odyssee im Weltraum“ ist eine emotionale Maschine. Sie wendet sich im Verlauf des Films gegen ihre Schöpfer und versucht ihre Abschaltung durch Mord zu verhindern. In den Filmen der Star Wars Saga hat C-3PO einen menschlichen Körperbau. Der goldene Droide hilft den Helden bei Übersetzungen in den Universen mit vielen verschiedenen Sprachen.

Die Verbindung zwischen Science Fiction in Büchern und Filmen und der technischen Welt war schon immer verblüffend eng. „Die Zukunft ist längst hier – sie ist halt nur

ungleichmäßig verteilt“, sagte William Gibson in einem Radiointerview [1] für den amerikanischen Sender NPR mit dem Titel: „The Science in Science Fiction“. Der Autor des Romans Neuromacer und Schöpfer des Begriffs „Cyberspace“ hat eine technisch interessierte Generation beeinflusst, von denen etliche führende Köpfe im Silicon Valley wurden. Der Begriff „Nerd“ dürfte auf viele der Science-Fiction-Drehbuchautoren zutreffen, genau wie es auf viele Entwickler in Start-ups und Technikkonzernen zutrifft. Eine gegenseitige Beeinflussung dieser beiden Welten ist daher durchaus plausibel.

Doch im praktischen Alltag mussten die Computernutzer auf ihrem Weg zu einem CUI erst einen gewaltigen Umweg gehen. Die Schnittstelle Mensch/Maschine bestand in den Anfängen aus Lochkarten. Später mussten endlose Befehle (Command Lines) über eine Tastatur in formalisierter Struktur eingegeben werden.

In der Entwicklung der Personal Computer (PC), brachte die grafische Benutzeroberfläche (Graphical User Interface) den Durchbruch zum Massenmarkt. Die Computersteuerung mit einer Maus wurde damals als großer Entwicklungsschritt gefeiert. Doch wenn man ihn rückblickend betrachtet, war das Erlernen der Hand-Auge-Koordination für die Mausbedienung ein Teil des Umweges. Die Bedienung mit einem oder mehreren Fingern auf einem berührungsempfindlichen Bildschirm (Touch) ist viel natürlicher. Doch das sollte erst mit dem Erfolg des Smartphones Einzug halten (siehe Abb. 1.3). Um dem Anwender den Umstieg von einer physikalischen in eine digitale Welt schmackhaft zu machen, hat man es mit Abbildern der Realität versucht. Fachleute bezeichnen diese Entwicklung im Design als Skeuomorphismus. Der erste Kalender auf dem Desktop sah aus wie ein Filofax. Selbst der Ledereinband wurde im Rechner nachempfunden. Gleicher gilt für Notizen. Die Anwendung Notizen sah lange aus wie der in den USA übliche gelbe Notizblock. Digitale E-Books legte man bei Apple anfänglich im iPhone und iPad in ein Holzregal. Der Papierkorb sah zu Beginn auf dem Desktop aus wie der Metalleimer in der

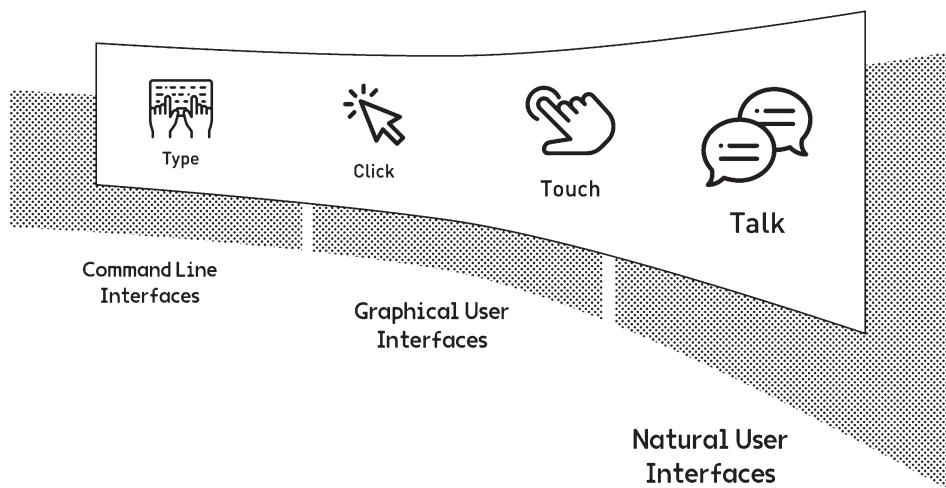


Abb. 1.3 Evolution von der Befehlszeile zur Spracheingabe

Ecke des Arbeitszimmers. Je ähnlicher die digitale Welt den bekannten Objekten ist, desto leichter fallen Akzeptanz und Nutzung, so das Kalkül der Designer. Dabei verwirren derartige Metaphern häufig, wenn beispielsweise das Symbol für die Sprachbox ein nachgebildetes Tonband ist oder für das Speichern eine Diskette. Derartige Objekte dürften den meisten Nutzern heute überhaupt nicht mehr bekannt sein.

Der Übergang zur Sprache (Natural User Interfaces) ist eine logische und zwangsläufige Entwicklungsstufe. Je leistungsfähiger Technologie wird, umso weniger müssen wir Menschen uns im Dialog verstellen. Mit anderen Menschen kommunizieren wir auch über Schrift- und gesprochene Sprache.

Wer glaubt, Conversational User Interfaces (CUI) sind eine Erfindung unserer Zeit, täuscht sich. Die Anfänge gehen zurück in die 1960er-Jahre. Der Informatiker Joseph Weizenbaum entwickelte 1966 das Programm ELIZA. Es ist die Simulation eines Dialogs zwischen Patienten und einem virtuellen Psychologen. In einem schriftlichen Dialog beantwortet der Nutzer Fragen oder macht Aussagen, auf die ELIZA mit Antworten reagiert. Das funktioniert über ein einfaches Parsingverfahren. Das Programm reagiert auf erkannte Schlüsselwörter und liefert eine vorab formulierte Antwort. Viele Nutzer gaben extrem unkritisch persönlichste Details im Dialog mit diesem frühen Chatbot preis. Weizenbaum war von der Offenheit der Probanden zunächst entsetzt. Der deutsch-amerikanische Informatiker leerte ab den 1970er-Jahren Computer Science am MIT in Boston. Zuvor war er am Aufbau vom Arpanet, dem militärischen Vorläufer des Internets, beteiligt. Die Erfahrung mit ELIZA trug dazu bei, dass Joseph Weizenbaum sich zu einem Kritiker gedankenloser Computerhörigkeit entwickelte.

Die Software ELIZA gilt auch als einer der ersten Anwendungsfälle für den Turing-Test. Den hatte der britische Mathematiker Alan Turing in den 1950er-Jahren entwickelt. Die Versuchsperson tippt ihre Aussage auf einer Computertastatur und erhält danach eine schriftliche Antwort. Vereinfacht gesagt, sollte der Test feststellen, ob eine Maschine mit dem Denkvermögen eines Menschen mithalten kann. Bei ELIZA konnten die Nutzer jedoch bereits nach wenigen Wortwechseln ahnen, dass sie mit einer Maschine kommunizieren und nicht mit einem, wie behauptet, menschlichen Psychologen. Der Turing-Test gewinnt in der Diskussion um künstliche Intelligenz, maschinelles Lernen und CUIs wieder an Popularität und wird derzeit immer wieder in den Medien zitiert.

1.2 Instant Messenger als Vorläufer heutiger Chatbots

Den ersten Vorläufern für (Chat)-Bots begegnet man bereits in einfachen Bulletin-Board-Systemen, der Vorform des Internets in den 80er-Jahren. Dann tauchten Chatbots in Instant-Messenger-Diensten wie ICQ (I seek you/Ich suche Dich) auf, der zeitweise zu AOL gehörte. In einigen Gruppen konnten Nutzer bei automatisierten Chatbots individualisiert einfache Informationen oder Grafiken abfragen und eigene Nachrichten in verschiedenen Chat-Räumen verbreiten. Der Erfolg dieser Messenger-Dienste auf mobilen Geräten in den 1990er-Jahren ist zum Teil mit dem relativ hohen Preis einer SMS über das Mobilfunknetz

zu erklären. Zum anderen war der direkte Dialog etwas Neues. Die schriftliche Kommunikation über Instant Messenger erfolgte anders als bislang gewohnt. Es muss nicht, wie bei der E-Mail, eine neue Nachricht vom Server abgerufen werden. Sie wird dem Empfänger direkt zugestellt. Die Nachrichten sind kürzer – jeder Gedanke bzw. jede Aussage ist eine separate Nachricht, was es einfacher macht, sich auf eine dieser Nachrichten zu beziehen, zu zitieren und darauf gesondert zu antworten. Die Dialoge sind einer gesprochenen Unterhaltung ähnlicher als schriftlicher Kommunikationen in Briefen oder E-Mails. Zugleich dokumentieren sie einen Gesprächsverlauf oft besser, als dies in der Abfolge von E-Mail-Dialogen möglich ist. In westlichen Ländern wurde die App WhatsApp sehr erfolgreich. So erfolgreich, dass Facebook sich gezwungen sah, das Unternehmen für einen hohen Betrag zu übernehmen, um die Konkurrenz zum eigenen Facebook Messenger zu verhindern. Vermutlich auch aufgrund der vielen negativen Nutzererfahrungen mit teilweise spammenden Chatbots in ICQ waren die Produktentwickler von WhatsApp immer sehr zurückhaltend in der Frage, ob und in welchem Umfang Chatbots in WhatsApp eine Rolle spielen können. Der Kurznachrichtendienst ist heute in der Hand von Facebook. Bis heute sind automatisierte Mechanismen auf WhatsApp nicht richtig erwünscht. Betreiber, die die Plattform als Newsletter-Distribution nutzen, werden behindert und teilweise gesperrt.

Ganz anders verlief die Entwicklung dagegen in China: Die chinesische App WeChat von Tencent hat die Nutzung eines Instant Messengers auf ein neues Level gehoben in dem neben Privatnutzern auch Unternehmen als Akteure in den Chats agieren. Inzwischen hat sich die App zum universellen Betriebssystem auf vielen Smartphones entwickelt. Die Nutzer chatten, shoppen, spielen und nutzen sie für ihre Navigation. Mit WeChat Pay kann man seine Einkäufe bezahlen. WeChat Pay fungiert als digitales Wallet, das mit einem Bankkonto verknüpft ist – ähnlich eines in die Messenger-App vollständig integrierten PayPal-Kontos. Neue Funktionen werden dem Nutzer innerhalb der App angezeigt und können direkt aktiviert werden. Für viele Funktionen müssen Nutzer nicht erst einen separaten App-Store besuchen und eine App installieren. Man muss sich keinen Namen einer App merken, um sie später installieren zu können. Es gibt in der Nutzung weniger Brüche und an vielen Stellen übernehmen Bots den Kundenservice, was stark zum Erfolg von WeChat als Universal-App beigetragen hat. Die große Verbreitung, mit über einer Milliarde aktiver Nutzer pro Monat, ist die beste Grundlage für ein CUI.

WeChat gilt weltweit als Vorbild für ein modernes, mobiles Ökosystem. Das sieht wohl auch Facebook-Chef Mark Zuckerberg so, der im März 2019 in einem vielfach beachteten Blogpost einen Strategieschwenk verkündete. Das verleitete die renommierte Online-Publikation „The Verge“ zur Überschrift „Facebook wants to be WeChat“ [2]. (Mehr zu WeChat findet sich in Kap. 7)

1.3 Das Voice Jail der Telefon-Hotlines

In den USA haben etliche Unternehmen bereits in den 1980er-Jahren in Call-Center-Automation investiert. Das Telefon als direkter Kommunikationskanal zu Kunden und Interessenten war schon in den 1970er-Jahren mit kostenfreien Hotlines etabliert, die

in Verbindung mit der alphanumerischen Touchtone-Nutzung offensiv im Marketing eingesetzt wurde („Call-1-800-PIZZA“). Hier mussten Nutzer Ziffern auf der Telefonastatur für ihre Auswahl drücken. Später konnten sie Nummern oder Menüpunkte einspielen. Entsprechend ausgerüstete Telefonanlagen wurden direkt bei den Anbietern oder ihren Call-Center-Dienstleistern fest installiert. Der Dialog-Flow war starr vorgegeben. In der Regel wurden die Auswahlmöglichkeiten dem Anrufer vorgelesen. Der musste sich alle Optionen merken und dann seine Auswahl in Form eines Tastenbefehls treffen. Etwas komplexere Vorgänge wurden sehr langwierig und waren für den Nutzer ein unerfreuliches Erlebnis. Das änderte sich kaum, als Nutzer ihre Auswahl nicht mehr nur per Tastendruck, sondern per Sprachbefehl eingeben konnten. Der Begriff von „Voice Jail“ [3] machte die Runde. Auch wenn entsprechende Systeme abschließend nicht positiv von Kunden wahrgenommen wurden, zeigten sie doch die Möglichkeiten auf, die durch automatisierte Kundenkommunikation möglich waren. So konnten Kunden, ohne Wartezeit zu jeder Tages- und Nachtzeit ihre Bedürfnisse ausführen, was vor Einführung des Internets eine Sensation war, aber selbst nach Einführung des Netzes ein echtes Serviceversprechen darstellte. Die US-Fastfoodkette Domino's hat dieses Know-how früh in seine App integriert. Seit Juni 2014 bietet das Unternehmen die Möglichkeit, per Sprachbefehl über das Smartphone Pizzen zu bestellen. Noch im selben Jahr startet Amazon sein sprachbasiertes Echo für Prime-Kunden.

1.4 Darum eignet sich Sprache für einen Dialog mit Maschinen

Alle Sprachen dieser Erde basieren auf einem klaren Regelwerk. Es gibt eine Grammatik, die Fälle, Zeiten und Konjugationen für Wörter vorschreibt. Gleichzeitig ist der Wortschatz einer Sprache endlich. Von daher können Maschinen Sprache gut analysieren. Sätze werden in Einzelteile zerlegt, Worte erkannt und ihre Beziehung zueinander analysiert und als Wahrscheinlichkeiten bewertet. Die Maschine an sich versteht Sprache nicht. Jedenfalls nicht so, wie wir Menschen es tun. Daher ist auch der Begriff „künstliche Intelligenz“ an dieser Stelle irreführend, denn mit Intelligenz, wie wir sie verstehen, hat die Spracherkennung noch nichts zu tun.

Das Natural Language Processing (NLP) beruht auf Algorithmen und maschinellem Lernen (siehe Abb. 1.4). Es geht um Statistik, um prozentuale Wahrscheinlichkeiten. Eine Maschine berechnet, zu welcher Wahrscheinlichkeit der Sprachbefehl beispielsweise der Wunsch nach einem Zugticket ist. Liegt die Wahrscheinlichkeit bei einem vom Programmierer festgelegten prozentualen Wert, wird ein Prozess angestoßen, bei dem das System über eine Schnittstelle Kontakt zum IT-System der Bahn aufnimmt und eine Fahrkarte kauft. Es liegt in der Hand der Entwickler, ab welcher Wahrscheinlichkeit, auch Confidence Level genannt, die Maschine eine Handlung umsetzt. Liegt die Wahrscheinlichkeit unterhalb des festgelegten Wertes, fragt der Assistent noch einmal nach.

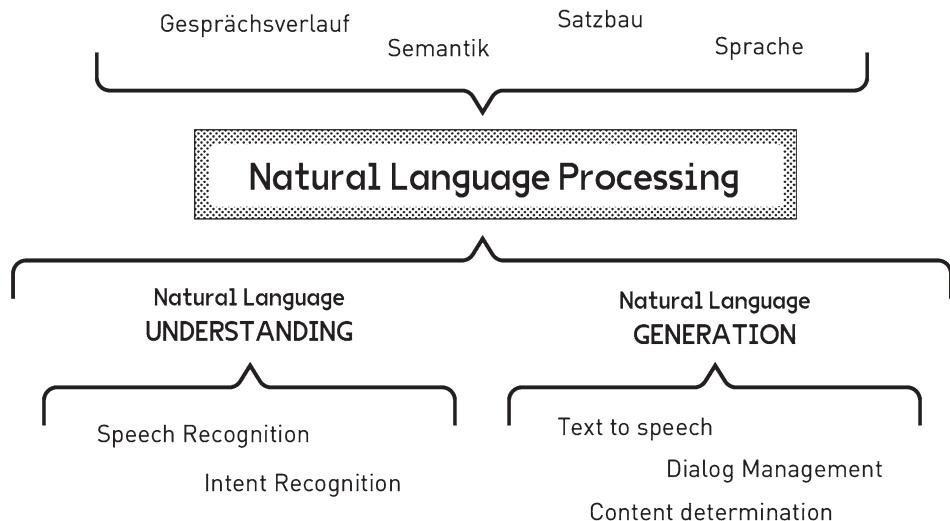


Abb. 1.4 NLP ist die maschinelle Form der Spracherkennung

1.5 Trainingsdaten in digitalisierter Form

Die Einfachheit spricht klar für natürlichsprachige Interfaces: Es müssen keine Befehle, keine bestimmten Formulierungen gelernt werden. Im Idealfall versteht die Maschine eine Anweisung auf Anhieb. Sprache ist die natürliche Form des Austausches zwischen Menschen.

Unter dem Oberbegriff künstliche Intelligenz (AI) verbergen sich im Zusammenhang mit Sprachverarbeitung in erster Linie Methoden des maschinellen Lernens. Damit macht das Interpretieren von Sprache einen Quantensprung. Maschinen lernen, wenn ihnen große Mengen an Trainingsdaten zur Verfügung stehen. Darin erkennen sie Muster, können Rückschlüsse ziehen und ihre Programme weiter verfeinern (siehe Abb. 1.5).

Es gibt einen nahezu unerschöpflichen Fundus an Trainingsdaten: Inzwischen ist fast jedes gedruckte Buch digital erfasst. Hinzu kommen tagesaktuelle Nachrichten auf Webseiten, Gesetzestexte, Verträge und jede Menge weiterer Texte auf Webseiten. Den großen Internetunternehmen wie Google, Facebook, Microsoft und Apple stehen darüber hinaus Milliarden E-Mails und Chat-Protokolle aus Messenger-Diensten zur Verfügung. Somit existiert eine grenzenlose Menge an digitalen Daten zur Mustererkennung von Sprache. Für maschinelles Lernen ist die Menge der vorhandenen Trainingsdaten entscheidend. Je größer die Grundlage, umso schneller und besser lernt das System.

Wie gut Maschinen natürliche Sprache bereits „verstehen“, zeigt die Entwicklung der Suchmaschinen in zwei Bereichen. Einerseits im Anwendungsgebiet von Sprachsuche, bei der Genauigkeit (Word Accuracy Rate) bei der Erkennung von gesprochenen Wörtern. Andererseits bei der inhaltlichen Interpretation von Suchanfragen: Anfänglich war es bei

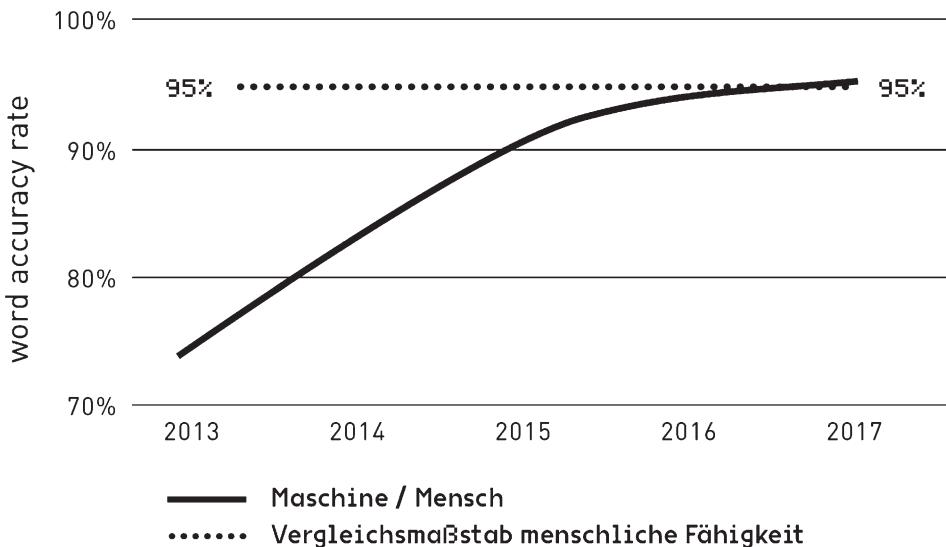


Abb. 1.5 Genauigkeit der Spracherkennung durch Maschinen

Suchmaschinen notwendig, Boolesche Operatoren (UND, ODER, NICHT) miteinander zu verknüpfen, um das Gesuchte zu finden. Da musste man ins Suchfenster eingeben: „Reifen UND wechseln NICHT Fahrrad“. Heute gibt man seinen Suchwunsch ein, als ob man einen Freund um Rat fragt: „Wie wechsle ich am schnellsten einen Reifen an meinem Auto?“ Die Systeme verstehen das Anliegen des Nutzers, den so genannten Intent, und liefern passende Antworten. Die Suchmaschine Google ist so ausgereift, dass einfache Antworten auf Faktenfragen direkt auf der Trefferseite beantwortet werden, ohne dass man den Link zur Zielseite anklicken muss. 8848 Meter lautet die Antwort direkt auf der Suchseite, wenn man die Frage „Wie hoch ist der Mount Everest?“ ins Suchfeld tippt.

Während bei der webbasierten Suche die Anzahl und Qualität der Links als auch die Häufigkeit der Aufrufe über die Relevanz der Inhalte entscheiden, geht es bei der semantischen Suche um das Verstehen der Nutzerfrage. Die NLP-Systeme zerlegen die Sätze, analysieren die einzelnen Bausteine und ihre Beziehung zueinander. Dabei werden die Satzelemente morphosyntaktisch ausgezeichnet. Das System weiß, dass „eingekauft“ auf die Grundform „kaufen“ zurückgeht. „Bei der semantischen Recherche kommen unter anderem auch Synonyme und Hyperonyme zum Einsatz. Zum Beispiel hat „PKW“ und „Automobil“ eine vollständige „synonyme“ Übereinstimmung. Dagegen hätte das Hyperonym „Fahrzeug“ eine 80-Prozent-Übereinstimmung mit der Frage nach einem PKW. So führt die semantische Recherche zu umfassenderen und präziseren Ergebnissen“ [4].

Das Natural Language Processing wird nicht nur bei CUIs eingesetzt, sondern spielt auch in weiteren Bereichen der maschinellen Sprachauswertung eine Rolle (siehe Abb. 1.6). Die Machine Translation ist eine automatische Übersetzung von einer in eine andere Sprache. Das wird von Übersetzungswebseiten oder -Apps genutzt. Das Information Retrieval

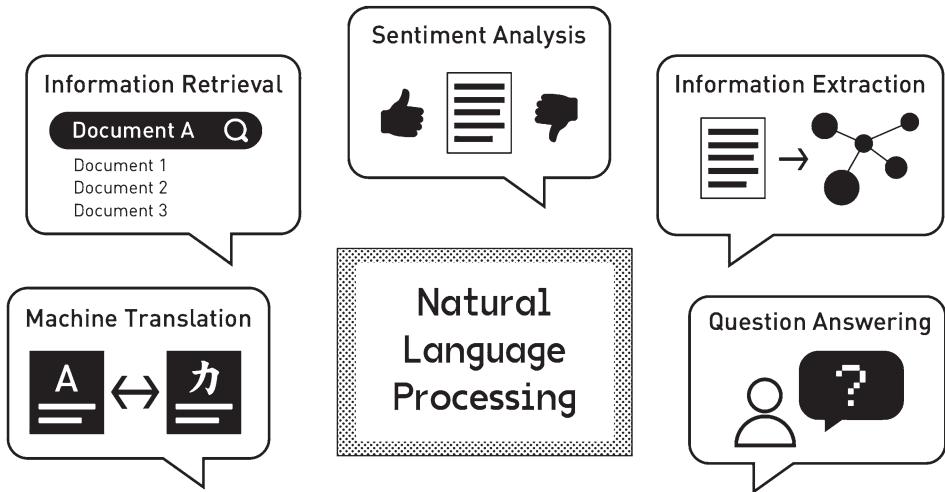


Abb. 1.6 Fünf Einsatzfelder von Natural Language Processing

beschreibt die Suche, beispielsweise nach Fakten und Auskünften. Das kann die Geschichte eines Unternehmens, die Höhe eines Berges oder die Erklärung der Fotosynthese sein. Somit findet es Anwendung in Suchmaschinen. Bei der Sentiment Analyse geht es um eine Bewertung der Gefühle des Absenders. Beantwortet werden soll die Frage: Ist der Autor des Beitrags/der Meinungsäußerung einer Sache gegenüber positiv oder negativ eingestellt? Die Technik hilft bei der maschinellen Auswertung von Bewertungen auf Webseiten und auch von Social-Media-Beiträgen. Die Information Extraction sucht in einer Vielzahl von Quellen nach den gewünschten Daten. Beispielsweise werden große Archive mit Lebensläufen von Bewerbern, Verträge oder Gesetzestexte durchsucht und analysiert. Das Question Answering ist die Suche nach Antworten auf klassische Fragen. Statt selbst eine FAQ-Liste zu durchsuchen, übernimmt die Maschine die Suche.

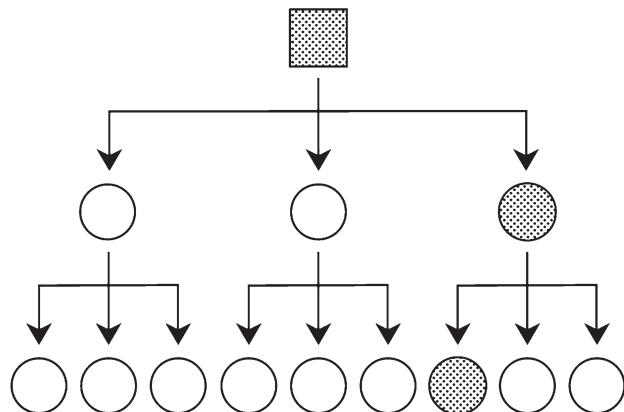
1.6 Das Versprechen von CUIs und intelligenten Assistenten

Die aktuell gebräuchlichen grafischen Benutzeroberflächen auf Computerbildschirmen und mobilen Geräten bieten viele Vorteile, wodurch sich Computersysteme und die selbstständige Bearbeitung komplexer Aufgaben durch Nutzer durchgesetzt haben. Zwei wichtige Probleme führen immer wieder zu Frustrationen bei den Nutzern, für die CUIs und intelligente Assistenten Lösungen anbieten: Hierarchie von Informationen und das Auswahlproblem (Paradox of Choice).

Kein Platz für Hierarchien

Jeder, der einmal auf der Website eines großen Unternehmens eine Produktinformation oder einen Dienst gesucht hat, kennt es: Vom Einstieg über die Homepage bis zum Formu-

Abb. 1.7 Relevante Informationen sind nicht immer einfach zu finden. Die sprachbasierte Suche überwindet das Hierarchieproblem grafischer Benutzeroberflächen



Ist für eine Schadensmeldung bei der Versicherung sind häufig vier, fünf oder sechs Klicks notwendig, bis man die gewünschte Information gefunden hat. Wenn es überhaupt gelingt, da man die Struktur der Hierarchie häufig gar nicht versteht und in einer Sackgasse landet. Dabei handelt es sich in den meisten Fällen nicht um das Versagen der Entwickler und Designer, sondern schlicht um ein Problem, das grafische Benutzeroberflächen strukturell in sich tragen: Sie bilden auf visuelle Weise eine Hierarchie ab, die in einem Großunternehmen mit vielen Aktivitätsfeldern und Zielgruppen eben tief und unübersichtlich ist. Diese Strukturen sind oft über Jahre gewachsen (siehe Abb. 1.7).

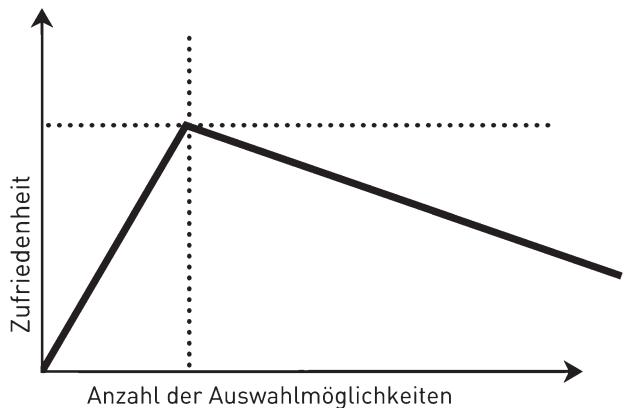
Das Versprechen von CUIs ist, dass der Nutzer einfach seinen Wunsch äußert und direkt und ohne weitere Umwege zu seinem Ziel gelangt. Durch natürliche Sprache und die Verfahren des NLP, sowie mithilfe von entsprechend vertagten Inhalten und Zielseiten gelingt es funktionalen Chatbots und Voice-Interfaces viele – in etlichen Fällen sogar alle – Hierarchiestufen zu überspringen.

1.7 Die Qual der Wahl – Finden statt suchen

Ein weiteres Paradoxon der modernen Welt wird als Paradox of Choice oder Auswahl dilemma bezeichnet (siehe Abb. 1.8). Zwar handelt es sich dabei nicht um ein originär digitales Problem, aber es wirkt in der digitalen Welt mit strukturell unbegrenztem Raum für Angebot und Auswahl besonders stark.

Ständen Sie schon einmal vor dem Regal mit Zahnpasta in einem großen Supermarkt? Falls ja, kennen Sie das Paradox of Choice. Lediglich zwischen zwei Zahncremes wählen zu können enttäuscht. Sie wünschen sich mehr Auswahl. Alles bis zu einem Dutzend Produkten steigert Ihre Zufriedenheit (happiness). Doch es gibt es einen Punkt, an dem Ihre Zufriedenheit mit steigender Auswahl nachlässt. Ab diesem Punkt überfordert Sie die Auswahl. Im Internet ist tendenziell jeder E-Commerce-Anbieter ein großer Supermarkt und jeder Nutzer kennt daher die Überforderung nicht nur bei Zahnpasta, sondern sämtlichen Produkt- und Servicekategorien.

Abb. 1.8 Weniger ist mehr:
Paradox of Choice – Die
Zufriedenheit nimmt ab, wenn
zu viele Dinge zur
Auswahl stehen



Hier ist ein CUI klar im Vorteil. Bei einer Handelsplattform sehen Sie bei einem CUI nicht, wie groß die Auswahl an Zahnpasta ist. Sie können nach Ihrer Lieblingsmarke fragen oder sich Produkte mit einer bestimmten Eigenschaft oder gewünschten Zutaten ansagen lassen. Ein CUI bietet im Vergleich zu einer klassischen Webseite noch einen weiteren Vorteil. Eine hierarchische und strukturierte Gliederung von Inhalten oder Produkten entfällt.

Bei einer Handelswebseite müssen alle Produkte bzw. Informationen in hierarchischen Kategorien eingesortiert werden. Hintergrund ist der begrenzte Platz auf einer Homepage, die auf einem Monitor aufgerufen wird. Auf den Bildschirmen der Smartphones und Tablets fällt dieses „Schaufenster“ noch kleiner aus. Zu einem CUI sagt der Nutzer ohne spezielle Befehle oder hierarchische Strukturen zu kennen zum Beispiel: „Ich suche eine Zahncreme ohne Mikro-Plastik und mit viel Fluorid“ oder „Ich brauche neues Klopapier.“ Bei letzterem Produkt bedarf es nicht viel Vertrauen, um es sozusagen „blind“ über ein CUI zu bestellen.

1.8 Herausforderungen von CUIs und intelligenten Assistenten

Bei anderen Produkten dürfte die fehlende Abbildung zunächst ein Hindernis darstellen. Bei dem Wunsch: „Ich hätte gern neue Sneakers für mich. In Grün und Größe 45.“ Hier sieht der Nutzer die Auswahl an Schuhen vorab nicht. Schließlich fehlt reinen Voice-Systemen der Bildschirm. Somit zeigt sich das Vertrauensdilemma besonders stark. Aber selbst bei Systemen mit Bildschirmen ist strukturell nur Raum für eine sehr eingegrenzte Auswahl vorhanden. Die entscheidende Frage lautet: Werden Nutzer bei CUIs quasi blind Produkte bestellen? Werden Nutzer den Intelligenten Assistenten tatsächlich als „ihren Assistenten“ verstehen?

Das Vertrauensdilemma oder auch Vertrauensparadox ist ein grundlegendes Phänomen der CUIs (siehe Abb. 1.9). Inwieweit wird der Nutzer einem intelligenten, virtuellen Assistenten mit persönlichen Informationen und Anliegen vertrauen? Wir beobachten eine

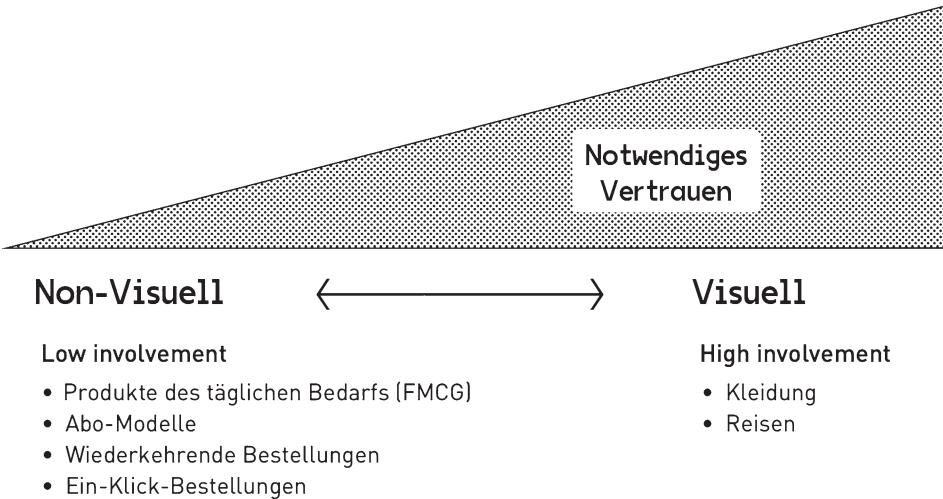


Abb. 1.9 Das notwendige Vertrauen variiert mit der emotionalen Bindung an das Produkt

Schere zwischen Vertrauen und Nutzungsintensität bereits bei Suchmaschinen wie Google und Social Networks wie Facebook. Eigentlich vertraut der Nutzer diesen Anbietern nicht wirklich, was immer wieder in unterschiedlichen Befragungen zu Tage tritt (Stichworte Datenkrake, Datensicherheit). Doch aus Bequemlichkeit oder weil der persönliche Nutzen so groß ist, greifen dennoch täglich Millionen Nutzer auf die genannten Angebote zurück. Dieses Vertrauensparadox werden wir aller Voraussicht nach auch bei CUIs erleben. Vielleicht sogar in noch viel stärkerer Ausprägung.

Wenn Maschinen autonom Entscheidungen treffen, die unser Leben eigentlich erleichtern, wie stark vertrauen wir den intelligenten Assistenten? Vermutlich werden sich schnell Antworten finden. Schließlich stellte sich die gleiche Frage in den Anfangstagen des E-Commerce bei Kleidung und insbesondere bei Schuhen. Die großzügigen Rücksendebedingungen der großen Anbieter haben das Vertrauensdilemma schnell überwunden. Das müssen auch CUIs schaffen – nicht nur im Versandhandel.

Eine weitere Herausforderung ist die schlechte Discoverability. Es ist schwer, dafür ein passendes deutsches Wort zu finden. Es beschreibt die Möglichkeit, gesehen zu werden oder neue Dinge zu entdecken, also die Auffindbarkeit. Grafische Benutzeroberflächen haben dieses Problem ebenfalls, weil nur ein Bruchteil des Angebots auf dem Bildschirm Platz findet.

Doch Websites haben Möglichkeiten gefunden, eine Vielzahl von Angeboten gleichzeitig oder wechselweise zu präsentieren. Einige davon prominent mit visuell ausgestatteten Teasern. Werbung ist eine Form davon, die die Aufmerksamkeit der Nutzer bewusst vom eigentlichen Trägerthema ablenkt und auf diese Weise auf neue Marken und Produkte aufmerksam macht, die der Nutzer unter Umständen noch nicht kannte und möglicherweise noch nicht einmal suchte. Die ganze Aufmerksamkeitsökonomie als Bestandteil von Waren- und Produktwelt basiert auf dieser Möglichkeit.

Bei CUIs ist dies nicht, oder nicht gut möglich. Einerseits ist der Startpunkt eines CUIs im Zweifel der einfache Texteingabeschlitz in einem Messenger oder ein Smartspeaker und dessen Mikrofon. Jedenfalls ist es kein Ort für ablenkende Botschaften. Die angestrebte Antwort eines CUIs oder intelligenten Assistenten darf nicht das Vorlesen einer endlosen Trefferliste sein. Im Idealfall ist es die einfache Ausführung eines Nutzerwunsches ohne weitere Nachfragen. Platz für ablenkende und ergänzende Botschaften ist hier zunächst nicht vorhanden.

Beispieldialog [5]:

- *Frage an Google Home:*

Ok Google, wie sieht mein heutiger Tag aus?

- *Lautsprecher:*

Es ist 7:44 Uhr. Der Tag wird sonnig. Es sind 21 Grad. Die Höchsttemperatur erreicht 37 Grad. Deine Fahrzeit mit öffentlichen Verkehrsmitteln zur Arbeit dauert aktuell 39 Minuten. Ganz nebenbei: Disney's Beauty and the Beast hat heute Premiere (Musik ist zu hören). Für mehr Film-Spaß, frag mich etwas über Bell. Einen schönen Tag noch. Hier kommen die Schlagzeilen ...

Die fehlende Sichtbarkeit in CUIs wird für Marken zu einer neuen Herausforderung. Das gilt insbesondere für Produkte mit einer niedrigen emotionalen Bindung (low emotional involvement). Auf die Bestellung „Ich brauche neue Batterien“ wird das CUI vielleicht noch nach dem Format, AA oder AAA, fragen, doch die Marke dürfte eine untergeordnete Rolle spielen. Vermutlich gibt Amazon hier seinen Basics-Produkten den Vorzug gegenüber anderen Batteriemarken. Für die Hersteller bedeutet das neue Investments in CUI-Marketing. Anbieter von CUIs und intelligenten Assistenten müssen nach Lösungen für die schlechte Discoverability und Findability von Marken suchen.

Für Nutzer stellt sich die Frage, woher und auf welche Weise erhalten sie Anregungen? Natürlich gibt es dabei Unterschiede zwischen so genannten High-Involvement-Produkten (Reisen, Auto usw.) und Low-Involvement-Produkten (Seife, Joghurt usw.). Anbieter von FMCG (Fast Moving Consumer Goods wie Lebensmittel, Deo etc.) müssen sich vollkommen neu auf CUI-Marketing einstellen. „Es wird zukünftig eine Herausforderung sein, nicht mehr nur in das Relevant Set von Menschen zu gelangen, sondern in das Relevant Set von Algorithmen. Alexa & Co könnten die „Markenkiller“ sein, die als Gatekeeper nur noch wenige und bevorzugt eigene Marken unterstützen. Gerade der habitualisierte Kauf von FMCG-Marken wird nicht mehr jedes Mal aufs Neue abgewogen, sondern durch eine sprachliche Anweisung vom Kunden oder durch den Algorithmus automatisch durchgeführt. Marken haben dann kaum mehr die Möglichkeit, Präferenzen auf klassische Art und Weise zu bilden“ [6].

Noch in einem anderen Bezug ist die schlechte Discoverability von CUIs bemerkenswert: Für viele Nutzer stellt das Nichtwissen, was man mit CUIs eigentlich tun kann, eine der größten Hürden dar: Das größte Problem haben Nutzer mit Alexa, weil sie nicht wis-

sen, was sie tatsächlich sagen und fragen können. Ein Grund, weshalb alle Anbieter beispielsweise ihren Smartspeakern kleine Bedienanleitungen mitgeben, auf denen einige grundsätzliche Sprachbefehle und beispielhafte Skills und Actions notiert sind. Nutzer erhalten regelmäßige Mails mit neuen Beispielen, wonach sie mit einem CUI suchen könnten. Gedruckte Gebrauchsanweisungen? Hat man seit der Einführung von Smartphones nicht mehr gesehen.

1.9 Multimodal könnte in vielen Anwendungsfällen die Lösung sein

Die Lösung für die angesprochenen Herausforderungen der Sichtbarkeit, des Vertrauens und dem Entdecken von Neuem (Discoverability) könnten multimodale Systeme sein. Gemeint ist damit ein Zusammenspiel zwischen Smartphones, Smartspeakern und Assistenten mit einem Bildschirm. Neben der Sprach- oder Texteingabe entwickeln sich in multimodalen Systemen Sensoren zum Auslöser von Aktivitäten. Angenommen die Smart-Home-App auf dem Smartphone registriert über das GPS-Signal, dass sich der Nutzer auf dem Heimweg befindet. Sie sendet diese Information an den App-Anbieter, der die Daten an den Thermostaten in der Wohnung weiter leitet. Dort wird die Heizleistung hochgefahren, so dass der Nutzer bei seiner Ankunft warme Wohnräume betritt. Gleiches kann natürlich auch durch Wearables erfolgen. Somit erleben wir eine API-sierung der Welt. Nur wenn bei allen erdenklichen Systemen passende Schnittstellen vorhanden sind, können Aufgaben umgesetzt werden.

In diesem ersten Kapitel haben wir eine Übersicht darüber gegeben was Conversational User Interfaces und intelligente Assistenten ausmacht, woher sie stammen, was diese für Nutzer versprechen. Das vermutlich beeindruckendste CUI und Assistenten-System, das der Öffentlichkeit bislang präsentiert wurde, dürfte Google Duplex sein. Diesen Service stellte Google-CEO Sundar Pichai auf der Entwicklerkonferenz I/O im Mai 2018 vor. Dabei absolvierte der Assistent auf verblüffende Weise in natürlicher Sprache einen Telefonanruf in einem Friseurladen, um eigenständig einen Termin zu vereinbaren. Die Maschine erhält von ihrem Nutzer einen Auftrag und setzt sich selbstständig in Bewegung, um die Aufgabe im Dialog mit einem echten Menschen auszuführen. Duplex ist quasi der digitale Golem.

Literatur

1. Gibson, Wiliam, The Science in Science Fiction (22.10.2018) <https://www.npr.org/2018/10/22/1067220/the-science-in-science-fiction?t=1569849740886>, zugegriffen: 21.11.2019
2. Statt, Nick; Liao, Shannon, 8.3.2019, <https://www.theverge.com/2019/3/8/18256226/facebook-wechat-messaging-zuckerberg-strategy>, zugegriffen: 21.11.2019

-
3. Urban Dictionary, voicemail jail, <https://www.urbandictionary.com/define.php?term=voicemail%20jail>, zugegriffen: 21.11.2019
 4. Buxmann, Peter, Schmidt, Holger: Künstliche Intelligenz – Mit Algorithmen zum wirtschaftlichen Erfolg, Springer Gabler (2019)
 5. Meunier Bryson, Twitter, „New Beauty & the Beast promo is one way Google could monetize Home“ (16.03.2018), <https://twitter.com/i/status/842358950536318976>, zugegriffen: 21.11.2019
 6. Rinsum, Helmut van: Uwe Munzinger: „Alexa könnte Markenkiller sein“ (18.02.2019), <https://ki-marketing.com/uwe-munzinger#more-430>, zugegriffen: 21.11.2019



Stand heute: Sechs Beispiele für funktionierende CUIs

2

Zusammenfassung

Bei Sprachassistenten denkt jeder sofort an den Google Assistant und Amazon Alexa. Die beiden US-Unternehmen dominieren aufgrund ihrer Größe, Verbreitung sowie der hohen Zahl an Geräten mit Sprachsteuerung die Wahrnehmung der Thematik. Doch es gibt weitere gute Beispiele, die zeigen, dass bereits heute natürliche Sprache von Maschinen verstanden und in Aktionen umgesetzt werden kann. Lernen Sie in diesem Kapitel sechs funktionierende CUIs kennen.

Bislang verfolgen CUIs stets ein eindeutiges und oft überschaubares Ziel. Die einen wollen den Kundenservice für ein Unternehmen verbessern, die anderen beantworten allgemeine, aber sehr konkrete Fragen. Die ultimative Herausforderung der Entwickler lautet jedoch: Mit einer Maschine sprechen, als ob es ein Mensch wäre. Genau das verfolgt die „Alexa Prize Socialbot Grand Challenge“ [1]. 2019 veranstaltete Amazon diesen Entwicklerwettbewerb bereits im dritten Jahr. Es gewinnt der Bot, mit dem sich Alexa-Nutzer am besten und längsten unterhalten. Es geht um „absichtslose“, also ziellose Konversation über das Wetter, die Lieblingsmusik oder die besten TV-Serien. Es geht um Small Talk, das schwierigste Feld der Welt, in dem es keine thematische Beschränkung, dafür aber viele Regeln gibt, die das ganze vergnüglich machen soll.

Amazon lobt zur Teilnahme Preisgelder von über einer Million US-Dollar aus. Im Jahr 2018 ging der Preis an das Team Gunrock von der University of California in Davis. Der Wettbewerb ist so eine Art „Mondlandung“ der CUI-Szene. Die Unterhaltungs-Bots sind in der Praxis irrelevant, doch während der Entwicklungsarbeit werden etliche Hindernisse überwunden und wertvolle Erkenntnisse gewonnen. Der Gewinner erhält 500.000 Dollar

Preisgeld. Zusätzlich wird ein Forschungsetat von einer Million Dollar ausgeschrieben für den Bot, mit dem sich Nutzer im Durchschnitt mindestens 20 Minuten unterhalten und wenigstens eine 4.0-Bewertung vergeben.

Die meisten intelligenten Assistenten werden nicht in derartig esoterischen Gefilden betrieben, sondern entstehen mit klaren Kommunikations- und Geschäftszielen, die Nutzern einen bestimmten Vorgang erleichtern. Die Wissensdomänen bzw. Anwendungsmöglichkeiten sind bei den meisten dieser Praxisfälle klar umrissen. Die Vielfalt und Komplexität an Spracheingaben und erwarteten Ausgaben ist deutlich reduziert im Vergleich zum Parlieren im allgemeinen Small Talk.

2.1 Google Assistant und Amazon Alexa

In der öffentlichen Wahrnehmung sind die beiden Beispiele von Amazon und Google für CUIs am bekanntesten (siehe Abb. 2.1 und 2.2). Schließlich gehören sie zur größten Suchmaschine der Welt bzw. dem größten Online-Händler. Entsprechend unterschiedlich sind die Herangehensweisen der beiden Anbieter. Google möchte mit seinem Assistenten das indexierte Wissen auch sprachbasiert zugänglich machen. Amazon möchte Produkte verkaufen.

Googles Assistent spielt seine Stärken in mobilen Anwendungen aus. Als Teil der Android-Smartphones hilft der Assistent, wenn man im Auto die Augen auf der Straße behalten muss oder man unterwegs die Hände voll hat und nicht tippen/wischen kann. Ohne hinzusehen, erhält man die Auskunft, wann der Flieger an welchem Gate geht oder wie morgen das Wetter an der Reisedestination aussieht.

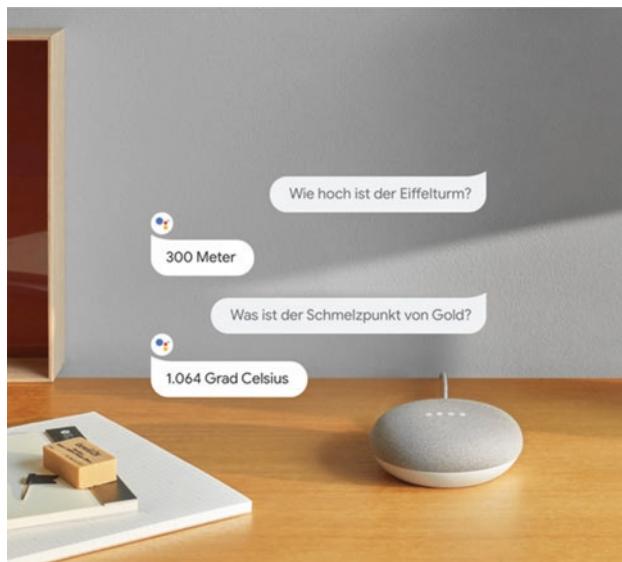


Abb. 2.1 Google Home Mini und Amazon Echo sind die bekanntesten CUIs im Consumer Segment. (Fotos: Google)

Amazons Alexa arbeitet ebenfalls an einer möglichst breiten Anwendungsbasis. Sie ist in Echo Dots, Echo Show, Fire TV Sticks sowie weiteren Amazon-Produkten integriert. Man findet inzwischen beide Assistenten in diversen Produkten von Drittanbietern. Dazu zählen unter anderem Autos, Lautsprecher von Sonos und Bose und mit kleinem Umweg auch die Symfonik-Lautsprecher sowie die Tradfri-Leuchten von Ikea. Überall beantworten der Google Assistant und Alexa Fragen oder schicken Produkte an die Wunschadresse des Anwenders.

Beide Anbieter setzen auf eine große Verbreitung. Für Alexa beispielsweise kann jeder einen Skill anbieten. Das führt zu einer umfangreichen Auswahl. Rein statistisch kommen jeden Tag in den USA 85 und in Deutschland 13 neue Skills hinzu. Zu Jahresbeginn 2019 lag die Zahl der verfügbaren Skills bei knapp unter 60.000 in den USA und knapp unter 8000 in Deutschland [2].

Die Entwicklung bei den Skills hat durchaus Parallelen zur frühen Entwicklung von Webseiten im Internet oder von Apps in App-Stores. Das Angebot reicht von simplen Funktionen, die einfach nur Meeresrauschen wiedergeben bis zu komplexen Smart Home Skills, mit denen sich Heizung, Licht und Musik im Haushalt steuern lassen. Hier sind ambitionierte Entwickler unterwegs, genau wie große Konzerne. Bei diversen Fluglinien kann man sich über Tarife und Verspätungen informieren. Die Deutsche Bahn sucht passende Verbindungen aus dem Fahrplan und verkauft auch das Ticket.



Abb. 2.2 Google Home Mini und Amazon Echo sind die bekanntesten CUIs im Consumer Segment. (Fotos: Amazon)

Bei der Vielzahl der Angebote, die man eigentlich nicht sehen kann, entsteht ein Problem mit der Sichtbarkeit. Welche Skills gibt es überhaupt? Wie finde ich neue, interessante Skills? Hier wird es zu einer Bereinigung des Angebots kommen. Bei Amazon muss sich ein Anbieter die Platzierung im impliziten Rank kaufen, genau wie man im Handel den Platz im Supermarktregal kauft. Bei Google entwickelt sich derzeit ein Algorithmus, der viel Ähnlichkeit mit dem Page Rank aufweist.

Neben Googles Assistenten und Amazons Alexa gibt es natürlich noch weitere CUIs wie Apples Siri, Microsofts Cortana und Samsungs Bixby. Auf sie gehe ich im Kap. 6 ein. Doch auch sie haben ihren Ursprung in einem Betriebssystem und weiten ihre Funktionen schrittweise auf andere Plattformen, Kanäle und Clients bzw. Geräte von Drittanbietern aus (Lautsprecher, Smart Home etc.). Die Verbreitung dieser CUIs dürfte in westlichen Ländern bzw. englischsprachigen Regionen am größten sein. Einen Blick nach Asien werfe ich in Kap. 7.

2.2 LH Group Chatbot: Mein Flug wurde gestrichen

Nach der erfolgreichen Einführung von Mildred im Facebook Messenger, einem Best-Price-Chatbot, startet die größte deutsche Fluggesellschaft einen zweiten Service Chatbot. Da der Name „LH Group Chatbot“ [3] etwas sperrig ist, stellt sich der Chatbot beim Leser als Elisa vor und macht deutlich, dass man es mit einem Computersystem zu tun hat (siehe Abb. 2.3). Der Zusatz „Group“ macht deutlich, dass der Chatbot für die gesamte Gruppe, bestehend aus Lufthansa, Swiss und Austrian Airlines, genutzt wird. Anfänglich versteht Elisa nur Englisch.

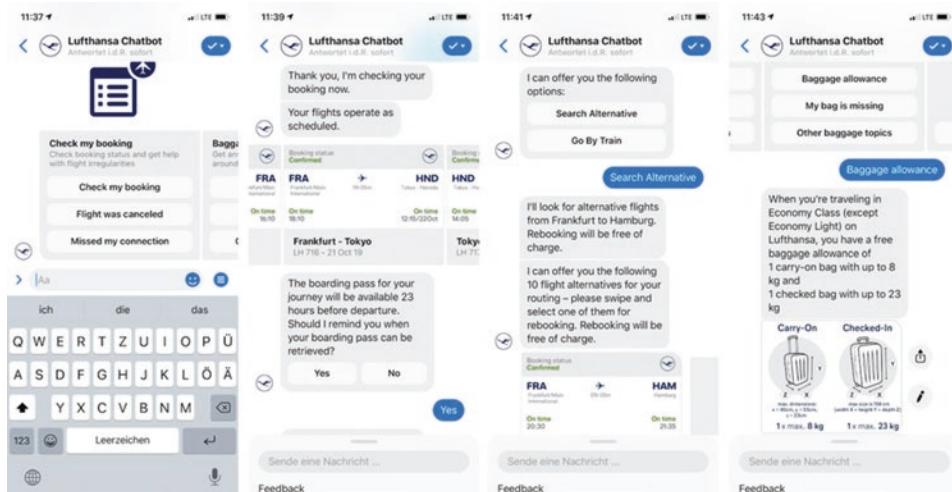


Abb. 2.3 Lufthansa Chatbot Elisa hilft dem Passagier während seiner Reise. (Foto: Lufthansa App/ Screenshot vom Autor)

Man kann dem Chatbot Feedback geben, also berichten, ob etwas besonders gut oder schlecht während einer Flugreise verlaufen ist. Elisa gibt Auskünfte über den Status des aktuellen Fluges, also beispielsweise, ob sich die Abflugzeit oder das Gate verändert haben. Man kann nach zugelassenem Gewicht sowie Maßen für die Koffer fragen oder verloren gegangenes Gepäck melden. Elisas Kernaufgabe sind jedoch so genannte Irregularities, also Unregelmäßigkeiten im Flugbetrieb. Wird ein Flug gestrichen, laufen mehrere hundert Passagiere, je nach Flugzeugtyp, direkt an die Schalter der Fluglinie im Abflugbereich. In der Regel gibt es nicht mehr als ein, zwei Mitarbeiter, die sich dann um Umbuchungen kümmern. Dabei wollen alle Betroffenen zum gleichen Ziel und die Umbuchung wiederholt sich dutzendfach. Diesen Vorgang automatisiert die Lufthansa, um zum einen ihr Personal zu entlasten, und zum anderen dem betroffenen Fluggast einen besseren Service zu bieten.

Bei einer Flugstreichung oder einem Ausfall identifiziert sich der Passagier über seinen Buchungscode oder die Ticketnummer sowie seinen Nachnamen im Messenger. Der Chatbot erkennt den betroffenen Flug und schlägt alternative Routen zum Ziel vor. Der Kunde wählt seinen Favoriten und wird automatisch auf den neuen Flug bzw. die Flüge, falls man noch mal umsteigen muss, gebucht. Eingechecktes Gepäck wird automatisch mit umgeleitet. Darüber erhält der Fluggast eine Information, genau wie seine neue, digitale Bordkarte. Im Idealfall erfolgt die Umbuchung innerhalb weniger Minuten und der Fluggast kann seine Reise fortsetzen. Kein Warten in telefonischen Hotlines oder der Schlange vor einem Schalter am Flughafen.

Die Airline erntet damit Low hanging fruits. Den Service für ihre Fluggäste muss sie ohnehin anbieten. Bei standardisierten Abläufen werden durch das CUI die Personalkosten gesenkt, denn 60 bis 70 Prozent der Fälle lassen sich automatisiert abwickeln. Der Service ist 24/7 und rund um den Globus verfügbar. Der normale Fluggast erhält mit dem Chatbot ein individuelles Serviceangebot, das bislang nur Statuskunden vorbehalten war.

2.3 Maybe: Lily bringt dem Zuhörer Chinesisch bei

Je enger der Funktionsumfang eines CUI definiert ist, desto besser funktioniert es für seine Nutzer. Der Sprach-Lern-Lautsprecher Lily ist dafür ein gutes Beispiel (siehe Abb. 2.4). Wobei man direkt einschränken muss: Seinen praktischen Nutzen hat Lily noch nicht unter Beweis gestellt. Es ist ein Projekt des US-Unternehmens Maybe und wird auf der Crowdsourcing-Plattform Indiegogo angeboten. Es ist also noch ein Pilotprojekt. Doch schließt man vom Interesse der Geldgeber auf die Eingangsthese, wird Lily ein großer Erfolg. Hinter Maybe stecken zwei erfahrene Toningenieure. Sie wollten 10.000 Dollar zusammen bekommen, um ihre Idee eines Chinesisch-Lern-Lautsprechers zu realisieren. Sie bekamen mehr als das Hundertfache. Über 8000 Menschen investierten 1,3 Millionen Dollar in das Projekt.

Seine Funktion ist schnell erklärt: Chinesisch lernen. Sein Äußeres trägt sicherlich auch zum Erfolg bei. Die bauchige Form des Lautsprechers erinnert an eine Comic-Figur, die einem gleich sympathisch ist. Mit Lily kann man im Büro oder Daheim ohne großen



Abb. 2.4 Lily ist ein Sprachen-Lern-Lautsprecher vom US-Unternehmen Maybe. (Foto: Maybe)

Aufwand ein paar Lektionen der Fremdsprache zwischen die sonstigen Dinge des Alltags schieben. Man muss in keine Schule, keinen Klassenraum und lässt keinen Lehrer warten. Dennoch ist es ein individuelles und vor allem akustisches Sprachtraining. Das ermöglicht ein individuelles Lerntempo sowie die Kontrolle der korrekten Aussprache. Mit Hochchinesisch bzw. Mandarin haben die Erfinder sich direkt eine der am schwersten zu erlernenden Sprachen ausgesucht. In dieser so genannten Tonsprache verändert die Tonhöhe bei der Aussprache eines Wortes seine Bedeutung. Im Chinesischen gibt es vier respektive fünf unterschiedliche Tonhöhen. So etwas kann man nur lernen, wenn man die richtige Aussprache hört und ein Lehrer direkt eingreift und geduldig korrigiert. Genau diese Aufgabe übernimmt Lily. Mit dem Lautsprecher kommt eine App. Nur so können die Schüler auch die Schriftzeichen lernen und Schreibübungen absolvieren. Bei Lily geht es nicht nur um Sprachlectionen. Die Nutzer können sich auch Geschichten vorlesen lassen oder mehr über die Kultur des Landes erfahren – natürlich auf Chinesisch. Kinder lernen spielerisch, in dem sie passende Bilder in der App antippen, nachdem Lily die jeweilige Vokabel ausgesprochen hat.

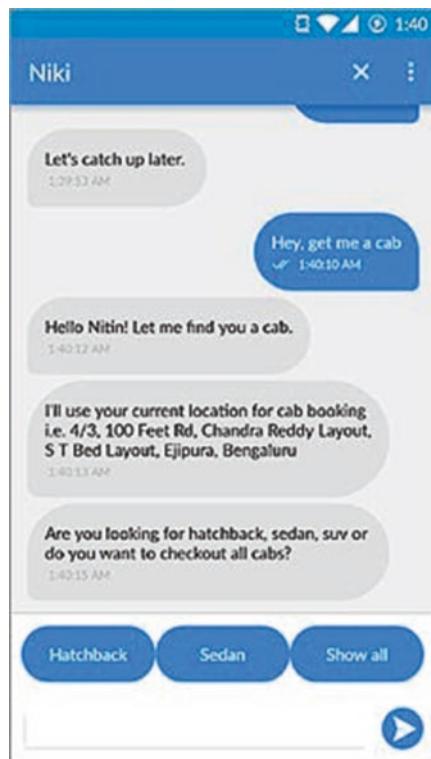
Hinter Maybe stehen die beiden Audio-Ingenieure Jason Wehner und Greg Timbers aus San Francisco. Wehner hat über 20 Jahre Berufserfahrung bei Unternehmen wie Harman, TDK und Vizio. Timbers war 43 Jahre lang beim Lautsprecherhersteller JBL tätig. Die beiden „Men of Sound“ legen bei Lily großen Wert auf guten Klang, wobei das bei einem Sprach-Lern-Lautsprecher sicher nicht die erste Erwartung der Nutzer ist.

Natürlich planen die beiden zukünftig Erweiterung auf andere Sprachen wie Französisch und Spanisch. Langfristig gedacht, kann die Technologie von Lily auch für (quasi) Simultanübersetzungen genutzt werden. Eine Person spricht den Satz auf Chinesisch, Lily gibt die Übersetzung auf Deutsch wieder [4].

2.4 Niki.ai: Eine Busfahrkarte kaufen

„Hello there, human!“, begrüßt der Chatbot den Nutzer. Niki macht dem Nutzer nicht vor, er spreche mit einem Menschen. Hier ist klar, ein Computer ist am Werk. Das indische Unternehmen ermöglicht es, Fernbus-Tickets über ein Voice-Interface zu kaufen oder ein Taxi zu bestellen (siehe Abb. 2.5). Man kann aber auch die Strom- und Wasserrechnung begleichen, die Prepaid-Karte des eigenen Handys aufladen und weitere Services mehr nutzen. Bemerkenswert ist dabei, dass diese Transaktionen, einschließlich der notwendigen Bezahlvorgänge, über ein Conversational User Interface vollständig durchgeführt werden. Eine Konsequenz, die zumindest in Europa und den USA, zumindest aktuell, noch nicht denkbar ist. Bemerkenswert ist ebenfalls, dass all dies nicht nur in Englisch, sondern in mehreren in Indien gebräuchlichen Landessprachen geschieht. Googles CEO Sundar Pichai – selbst indischer Abstammung – spricht von der „nächsten Milliarde Internetnutzer“ [5] und zielt damit insbesondere auf künftige Nutzer, die nicht mit einem PC oder Laptop aufgewachsen sind. Nutzer, für die grafische Benutzeroberflächen fremd sind, und die in vielen Fällen nicht oder schlecht Englisch sprechen. Von dieser Art Nutzern gibt

Abb. 2.5 Taxi-Bestellung mit Niki.ai. (Foto: Niki.ai)



es in Indien viele und genau für sie ist Niki.ai ein echter Segen, wenn sie auf Hindi, Tamil oder Bengali die Aufgaben des täglichen Lebens meistern können, indem sie einfach äußern, was sie wollen. Das Unternehmen arbeitet mit mehr als 50 Partnern zusammen, deren Produkte und Dienstleistungen mittels Niki.ai ausgewählt und gekauft werden können. Es wurden bereits über mehrere hundert Millionen Transaktionen abgewickelt. Die sogenannte Conversion-Rate, also der Prozentsatz begonnener Dialoge, die auch zum erfolgreichen Abschluss kommen, liegt bei über 92 Prozent und damit deutlich über den Branchenstandards westlicher Anbieter, die häufig nicht über 65 Prozent hinauskommen.

Niki.ai kann dabei mit gesprochenen und eingetippten Nachrichten umgehen und antworten.

Sachin Jasiwal, ein Absolvent des Indian Institute of Technology Kharagpur, hat 2015 mit einigen Kommilitonen in Bengaluru begonnen. Als Techbins Solutions Pvt. Ltd haben sie zunächst einen Chatbot (OnChat) im Facebook Messenger für die HDFC Bank umgesetzt. Es folgten schnell Kunden wie die Oxigen Wallet, Intex Smartphones und Ticketgoose, eine der größten indischen Plattformen für Dienstleistungen wie Busfahrten, Taxi und Hotel.

Zu den ersten Investoren (Series A) von Niki.ai zählen Unilazer Ventures und SAP.iO. Später kam Ratan Tata als Investor hinzu. Er ist CEO der Tata Sons Ltd. Die Holding vereint Beteiligungen in der Auto- und Metallindustrie, Telekommunikation sowie Softwareentwicklung. Das Engagement dieser prominenten Investoren spricht für die Zukunftsaussichten von Niki.ai.

2.5 Nomi: Wo soll die Fahrt hingehen?

Prost Nomi! Der digitale Assistent in den Elektroautos der chinesischen Marke Nio genehmigt sich an einem heißen Tag einen virtuellen Drink. Auf dem kleinen, runden Display sind ein Glas und zwei Augen zu sehen. Nomi lebt in einer Kugel, die mittig auf dem Armaturenbrett sitzt (siehe Abb. 2.6). Die Kugel dreht sich in die Richtung der Person, die in dem siebensitzigen Fahrzeug gerade mit Nomi spricht. Das verstärkt den Eindruck einer Kopfbewegung. Die menschlichen Züge auf dem Display sind auf die Augen begrenzt. Gelegentlich sieht man Finger für ein Daumen-hoch-Signal oder gerötete Wangen, wenn dem Assistenten etwas peinlich ist. Es ist aber niemals ein Mund zu sehen. „Der ganze Eindruck wäre zerstört, sobald Sprachausgabe und Lippenbewegung nicht mehr synchron sind“, begründet Kaveh Shirdel, Senior Director UI/UX bei Nio die Gestaltung. Sein Designteam wollte die Akzeptanzlücke, wie der Begriff „Uncanny Valley“ auch übersetzt wird, vermeiden. Je realistischer eine künstliche Figur dem Menschen nachempfunden wird, desto eher kommt der Punkt, an dem sich Anwender vor der menschlichen Nachbildung gruseln.

Mit „Hey Nomi“ aktiviert der Anwender den Assistenten. Auf Wunsch kann man Nomi einen anderen Namen geben und diesen als Signalwort verwenden. Jeder Passagier im Fahrzeug kann sein eigenes Profil anlegen. So ist eine Unterscheidung bei



Abb. 2.6 Nomi spielt Gitarre, sobald im Nio Musik erklingt. (Foto: Dirk Kunde)

Musikwiedergabelisten oder persönlichen Vorlieben möglich. „Ich kann sagen, mir ist es zu warm“, so Shirdel. Nomi reduziert die Heizleistung, beispielsweise auf 21 Grad Celsius. Wiederholt der Anwender den Satz, weiß Nomi, dass diese Person es gern unterhalb von 21 Grad mag. Die Nomi-Funktionen sind eng an das Fahrzeug gebunden. Sprachbefehle steuern die Navigation, Entertainment (Musik, Telefonate, ein Foto schießen), den Komfort (Lüftung, Klimaanlage) sowie den Kontakt zum Nio Service Network. Man kann am so genannten Nio House einen Batteriewechsel oder eine Innenreinigung des Fahrzeugs vorbestellen. Nio bietet bislang mit dem ES 6 und ES 8 zwei batterie-elektrische SUV in China an. Mit dem Nio ET gibt es den Entwurf einer Limousine und mit Nio Eve die Studie eines autonom fahrenden Fahrzeugs. Nomi ist seit einem Jahr aktiv und soll Teil aller kommenden Modelle sein. Da das Unternehmen von Gründer William Li bislang nur in China aktiv ist, versteht Nomi nur Mandarin. Sobald die Fahrzeuge nach Nordamerika und Europa kommen, sind natürlich Spracherweiterungen geplant.

Nomi schlägt seinen Anwendern mit jedem Software-Update für das E-Auto neue Funktionen bzw. Sprachbefehle vor. Ohne konkrete Zahlen zu nennen, sagt Shirdel, dass die Akzeptanz des digitalen Assistenten im Auto extrem hoch sei: „Für die Chinesen ist es schon lange normal, eine Sprachnachricht ins Handy zu sprechen, anstatt einen Text zu tippen. Die Adaption sprachbasierter Assistenten war da nur noch ein kleiner Schritt.“ Noch gibt es keine Nomi für das Wohnzimmer, um beispielsweise den Ladezustand abzufragen oder Navi-Ziele vorab ans Auto zu übertragen. „Doch es ist eine evolutionäre Weiterentwicklung in iterativen Schritten“, sagt Shirdel. Er will nichts ausschließen. In der dazugehörigen Nio-App für Smartphones spielt der Charakter Nomi bereits eine Rolle, wenn auch noch nicht sprachbasiert.

2.6 Param: Den passenden Bewerber finden

Dass Conversational User Interfaces und sprachbasierte Assistenten nicht nur in Anwendungen für Endkonsumenten, sondern auch im professionellen Umfeld von Unternehmen und Institutionen gute Einsatzgebiete haben, zeigt der Fall Param auf interessante Weise. Große Unternehmen besetzen pro Jahr hunderte oder tausende Stellen mit den passenden Bewerbern. Das Sichten und Auswählen der eingehenden Bewerbungen ist ein aufwändiger Prozess. Mitunter fällt der nicht immer konsistent aus. Menschen fällen je nach persönlicher Stimmung unterschiedliche Entscheidungen und es gibt Wechsel innerhalb der HR-Abteilung. Um den Auswahlprozess zu vereinheitlichen, aber vor allem zu beschleunigen, gibt es Assistenten wie Param.ai. Das Unternehmen mit Sitz im indischen Hyderabad geht die Herausforderung mit einem Deep-Learning-Algorithmus an. Das System wertet offene Stellenbeschreibungen anhand von vergleichbaren Stellenbesetzungen aus der Vergangenheit ab und nutzt dabei die Möglichkeiten, mittels Natural Language Processing Stellenbeschreibungen und CVs auszuwerten. Welche Eigenschaften muss der Bewerber mitbringen? Mit wem wurde die Stelle letztendlich besetzt und war das erfolgreich? Daraufhin gleicht Param die Angaben der eingehenden Bewerbungen mit den offenen Stellen ab. Das alles erfolgt innerhalb weniger Sekunden für Dutzende Stellen. Das Ergebnis ist eine Vorschlagsliste mit Kandidaten, die zu einem Vorstellungsgespräch eingeladen werden.

Die Vorauswahl von Bewerbern erfolgt mithilfe eines Chatbots. Param.ai bietet mit Lisa ein passendes CUI an. Kandidaten beantworten hier Fragen nach ihrer Ausbildung, ihrem beruflichen Werdegang, Verfügbarkeit und Gehaltsvorstellungen. Das System vergleicht dabei auch die im Lebenslauf getätigten Angaben, mit den Anforderungen der offenen Stelle. Zudem wird im Dialog ein Softwareentwickler beispielsweise nach den letzten Tools gefragt, mit denen er programmiert hat. Stimmen die Angaben nicht mit den Angaben aus dem Lebenslauf überein, wird das notiert. Es hat natürlich Auswirkungen auf das Ranking des Bewerbers. HR-Manager erhalten eine anhand des Rankings zusammengestellte Auswahl geeigneter Kandidaten, die eingeladen werden sollten. Das CUI vereinbart auf Wunsch auch direkt im Dialog mit dem Kandidaten einen Termin zum Interview.

Gerade B-to-B-Anwendungen sind für den Einsatz von Conversational User Interfaces gut geeignet, da im professionellen Umfeld die so genannte Wissensdomäne beschränkt und die Erwartungshaltung seitens der Anwender und Nutzer klar definiert sind. Im Fall von Param wird kein Nutzer den Chatbot bitten, einen Witz zu erzählen und die Grenzen seiner Funktionalität unmittelbar akzeptieren.

Literatur

1. Alexa Prize, <https://developer.amazon.com/alexaprize>, zugegriffen 21.11.2019
2. Kinsella, Bret: „Amazon Alexa Skill Counts Rise Rapidly in the U.S., U.K., Germany, France, Japan, Canada, and Australia“ (02.01.2019), <https://voicebot.ai/2019/01/02/amazon-alexa-skill-counts-rise-rapidly-in-the-u-s-u-k-germany-france-japan-canada-and-australia/>, zugegriffen 22.11.2019

3. Der Facebook Chatbot von Lufthansa, <https://www.lufthansa.com/xx/de/chatbot>, zugegriffen 22.11.2019
4. Indiegogo.com: Lily: The 1st smart speaker that teaches Chinese, (28.01.2019) <https://www.indiegogo.com/projects/lily-the-1st-smart-speaker-that-teaches-chinese#/>, zugegriffen 22.11.2019
5. Bohn, Dieter: Chasing the next billion with Sundar Pichai, (29.05.2015) <https://www.theverge.com/a/sundars-google/sundar-pichai-interview-google-io-2015>, zugegriffen 22.11.2019



Woraus setzen sich CUIs und intelligente Assistenten zusammen?

3

Zusammenfassung

In diesem Kapitel erfahren Sie, worauf Sie bei der Planung und dem Aufbau eines CUIs oder Assistenten achten müssen. Zunächst geht es um Fragen wie: Hat der Assistent einen Namen und ein Gesicht? Gibt es ein Logo und wie soll die Stimme klingen? Danach geht es um technische Anforderungen: In welchem Umfang werden Eingabemöglichkeiten natürlicher Sprache in Form von gesprochenen Worten oder geschriebenen Sätzen angeboten? Aber auch Sensoren (IoT) können Inputs für CUIs liefern. Die Erkennung und Analyse des Nutzerwunsches (Intent) ist der erste Schritt. Aber auch die Anbindung von Business-Logiken mit externen Services gilt es zu berücksichtigen. Nach Verarbeitung und Erkennung von Befehlen muss die Umsetzung des Wunsches bzw. die Präsentation der Ergebnisse auf geeignetem Weg zurück zum Nutzer

Ein CUI ist ein komplexes System, das aus diversen Bausteinen besteht. Durch die Nutzung von natürlicher Sprache gibt es mehrere Wege bzw. Geräte für die Eingabe. Es folgen verschiedene Methoden zu Analyse und Auswertung des Inputs. Die Anbindung weiterer Business-Logiken und Services bestimmt über die weitere Verarbeitung. Die Antwort an den Nutzer bzw. die Umsetzung seines Wunsches sollte auf gleichem bzw. gewünschtem Wege erfolgen.

Abb. 3.1 zeigt schematisch den Ablauf eines Nutzer-Inputs/Outputs. In einem multimodalen System hat der Anwender die Wahl zwischen einer Texteingabe, beispielsweise auf seinem Smartphone und der Spracheingabe, beispielsweise über einen vernetzten Lautsprecher. Die dritte Möglichkeit sind weitere Maschinen. Natürlich können auch Sensoren Befehle an ein CUI senden. Im Bereich Smart Home könnte der Input vom Smartphone kommen, dass nur noch einen Kilometer von der Wohnung entfernt ist. Es hat einen

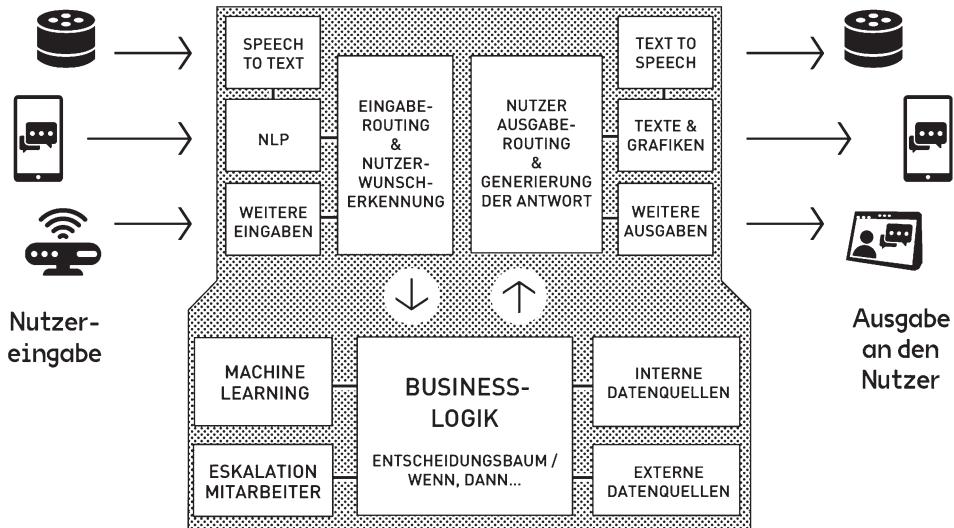


Abb. 3.1 Input, Verarbeitung und Output von Befehlen in CUIs

Geofence, der rund um den Wohnort gelegt wurde, „betreten“ und somit wird die Heizung vom System hochgestellt. Das Licht im Flur schaltet sich ein, sobald das Smartphone in direkter Nähe ist.

Wird ein Befehl vom Nutzer gesprochen, wandelt ein Speech-to-Text-Verfahren die Audiosignale in Text. Nun liegt in beiden Fällen der Nutzerwunsch in Textform vor. Ein NLP-Verfahren zerlegt, analysiert und bewertet die Eingabe. An dieser Stelle geht es darum, den Intent, also die Nutzerabsicht, korrekt zu erkennen. Will der Anwender ein Lied abspielen, Schuhe kaufen oder im Schlafzimmer das Licht anmachen? Zu welchem Prozentsatz ist das System sicher, den Nutzerwunsch korrekt erkannt zu haben? Der Entwickler legt hier einen Schwellwert fest. Überschreitet die Auswertung diesen prozentualen Wert, wird der Befehl weitergeleitet. Um beim Smart-Home-Beispiel zu bleiben, erfolgt hier beispielsweise ab 60 Prozent eine interne Anbindung an den Router oder den Hub des jeweiligen Systems (internal Data Sources). Bei komplexeren Wünschen, also dem Kauf von Schuhen oder der Buchung einer Reise, erfolgt die Weitergabe über Schnittstellen an externe Systeme (external Data Sources).

Dialog

- Nutzer: „Buch mir ein Hotelzimmer zur XY-Messe.“

Dieser Befehl erfordert im Idealfall die Anbindung diverser Systeme

- Maps: Welche Hotels liegen in Laufweite des Messegeländes?
- Kalender: An welchen Tagen ist die Messe eingetragen?

- Hotelzimmer: Zu welchen Buchungsplattformen bestehen Anbindungen? Gibt es ein präferierten Anbieter für den bereits ein Nutzerkonto hinterlegt ist? Gibt es hier Vorlieben für eine bestimmte Hotelkette oder Zimmerkategorie?
- Anreise: Sollen Fahrkarten für die Anreise erworben werden?
- Bezahlung: Sind Daten einer Kreditkarte oder eines anderen Bezahlsystems hinterlegt?
- Service: Bei kurzfristigen Buchungen oder kurz vor der Reise liefert das System aktuelle Verkehrsmeldungen, bekannte Störungen und die Wettervorhersage.

Die Business-Logik lernt mit jeder Buchung die Präferenzen besser kennen. Über Verfahren des maschinellen Lernens werden bisherige Reise ausgewertet und verbessern zukünftige Buchungen.

Sollte der prozentuale Schwellwert bei der Auswertung des Nutzerwunsches nicht erreicht werden oder sich im weiteren Verlauf Probleme mit externen Schnittstellen ergeben, muss es eine Fallback-Lösung geben. Im besten Fall ist das die Übergabe einer Buchung oder eines Einkaufs an einen Menschen im Callcenter des Anbieters.

Wurde der Wunsch korrekt umgesetzt, erfolgt die Ausgabe (Output). In den meisten Fällen werden Input- und Output-Kanal identisch sein. Spricht ein Nutzer seinen Wunsch, dann erwartet er oder sie in der Regel auch eine gesprochene Antwort vom System. Hier erfolgt eine Text-to-Speech-Umsetzung. Beim Beispiel der Hotelbuchung während einer Messe, könnte die Anzeige von in Frage kommenden Hotels auf einem Bildschirm hilfreich sein. Der Nutzer sieht auf einem Gerät mit Display eine Karte mit den Markierungen der verfügbaren Hotels bzw. die Ausstattung der Hotelzimmer. Per Fingertipp oder erneutem Sprachbefehl trifft er eine Auswahl. Hier können multimodale CUIs ihre Stärken ausspielen.

Die beste Spracherkennung nutzt nichts, wenn gesuchte Daten nicht performant zur Verfügung stehen und vom CUI genutzt werden können. Datenanbindung und Datenübertragungsgeschwindigkeit sind zwei wichtige Aspekte in der Nutzung. Die Akzeptanz von CUIs wird steigen, wenn man mit ihnen nützliche Dinge in angemessener Zeit erledigen kann. Je schneller ein CUI reagiert, desto öfter wird man es verwenden. Zum Beispiel die Organisation seines Tages im eigenen Kalender (siehe Cortana-Demo [1]). Langfristig wird Multimodalität mehr umfassen müssen als die Kombination von Sprache und Bildschirm. Dabei geht es um Gesten- und Bilderkennung, biometrische und Sensordaten. Darüber hinaus wird ein CUI in der Lage sein, aus der Sprache auch die Gefühlslage des Nutzers zu erkennen. Hand- und Augenbewegungen sowie Körpersprache werden von den Systemen erfasst und analysiert. Diverse Sensoren liefern weitere Daten. Das beginnt bei biometrischen Daten von Wearables bis hin zu Kameras aus dem Bereich Smart Home. Die Kombination sämtlicher Daten ermöglicht ein dezidiertes Bild der aktuellen Gefühlslage des Nutzers.

Dabei geht es nicht nur um die Verknüpfung unterschiedlicher Inputs. Entscheidend dürfte auch die Erkennung von Kontext, also Zusammenhängen sein. Was sagen biometrische Daten und ein bestimmter Gesichtsausdruck aus? Sind die Rückschlüsse des Systems korrekt? Dieser gesamte Prozess liegt im Bereich der Artificial Intelligence. Lernt die

Maschine aus Situation A genug, um Rückschlüsse auf Situation B zu ziehen? Wie vergleichbar sind menschliche Ausdrücke in der Kommunikation (Gesten, Stimmlage, Gesichtsausdruck) von einer zur anderen Situation? Die Herausforderungen für CUIs bestehen darin, die Feinheiten menschlicher Kommunikation zu erkennen und richtig zu deuten. Im Gespräch werden wir plötzlich lauter und sprechen schneller, wenn uns ein Thema am Herzen liegt oder wenn wir wütend sind. Gesprächspartner fallen sich ins Wort oder bestätigen Aussagen mit Nicken oder zustimmenden Worten. Der Gegenüber stellt Zwischen- und Verständnisfragen. Noch sind diese Feinheiten der menschlichen Kommunikation große Herausforderungen für CUIs.

3.1 Wo findet nun Artificial Intelligence statt?

Unter Artificial Intelligence versteht man eine Vielzahl unterschiedlicher Verfahrensweisen. Im Zusammenhang mit der Verarbeitung natürlicher Sprache ist vor allem das Machine Learning bedeutsam, Geräusche und Stimmen voneinander zu trennen und den Sound der Stimme in notierte Sprache – Buchstaben und Worte – zu wandeln funktioniert nur, wenn man unzählige Beispiel-Soundfiles einsetzt, um Computern zu „lehren“, welcher Laut welchen Buchstaben bedeutet. Auch die Analyse von niedergeschriebenen Nutzeräußerungen erfordert unzählige Textbeispiele, um die Algorithmen zu trainieren. Dazu mehr im folgenden Kapitel. Machine Learning ist die Grundlage für die Verarbeitung aller denkbaren multimodalen Inputs und ein Routing der Anfrage, zu den hinterlegten Businessprozessen und natürlich auch auf dem Rückweg der Antwort zum Nutzer. Im Idealfall erfolgen diese Schritte flexibel auf die Nutzerbedürfnisse angepasst und im Laufe der Nutzung und Zeit auch selbstlernend.

3.2 Was muss ich bei Planung und Aufbau eines CUI oder Assistenten beachten?

Im Folgenden werden exemplarisch einige der Fragestellungen aufgeführt, die man vor Erarbeitung eines CUI, Bots oder intelligenten Assistenten durchdenken sollte.

3.2.1 Branding, Naming und Gestalt

Wie soll der Assistent heißen? Der Eigenname eines Assistenten kann neutral beschreibend sein oder eher emotional. Letzteres führt auch direkt zu der Frage: Welche Persönlichkeit soll der Assistent haben? Welches Angebot bietet der Assistent dem Nutzer? Wer ist der Anbieter des Assistenten? Ist der Eigenname gleichzeitig das Aktivierungswort? In diesem Fall sind zwei Silben besser für die Erkennbarkeit als eine. Ein generischer Name wie „Assistent“ oder „Bot“ ist schlecht, weil er auch in Anfragen vorkommen könnte. Muss der Name in mehreren Sprachen funktionieren?

Zu den Überlegungen des Namens und einer Persönlichkeit gehört zwangsläufig auch die Gestaltung eines Visuals in Form eines Logos oder Icons. Die bildliche Darstellung des Assistenten verrät dem Nutzer bereits viel über seine Persönlichkeit. Immer wieder kommt in diesem Zusammenhang die Frage auf, ob ein Assistent einen Avatar benötigt. Eine Stimme vermittelt zwar einen Eindruck vom Charakter des Assistenten, doch eine visuelle Repräsentanz kann die Eigenschaften unter Umständen schneller und auch einfacher vermitteln. Wenn Sie sich für einen Avatar entscheiden, benötigt der Assistent dann sogar einen Körper? Mit dieser Frage berührt man das verwandte Gebiet der Roboter.

Zunächst wollen wir bei den auf Bildschirmen dargestellten visuellen Repräsentationen bleiben. Während Google Assistant, Amazon Alexa und Siri ohne derartige visuelle Darstellung auskommen, nutzt beispielsweise der chinesische E-Auto-Hersteller Nio bei seinem im Fahrzeug eingebauten Assistenten eine in ihrer Einfachheit an die frühen Smileys erinnernde Darstellung (siehe Kap. 2). Ja nach Situation transportieren nur wenige Striche die Mimik und eröffnen damit in der Kommunikation mit dem Nutzer eine interessante zusätzliche Ebene. Auf vollkommen andere Weise, letztlich aber mit vergleichbarer Methodik agiert Elliq [2] (siehe Abb. 3.2). Es ist eine Kombination aus Hard- und Software, die im Bereich der Altenbetreuung angesiedelt ist. Durch einfachste Mimik des Assistenten soll die Nutzung durch die vorwiegend wenig technikaffine Gruppe der Senioren erleichtert und angeregt werden.

Einen anderen Ansatz der visuellen Repräsentanz von Assistenten verfolgt das amerikanische Unternehmen Magic Leap. Das aus Florida stammende Unternehmen ist den meisten eher als Anbieter von visuell aufwendigen Augmented-Reality-Spielen bekannt und dadurch in die Öffentlichkeit gekommen, dass es über Jahre und unter Einsatz von Milliarden US-Dollar der Investoren hinter verschlossenen Türen an Entertainment-Lösungen arbeitete. Da der Durchbruch in diesem Feld offenbar nicht schnell genug



Abb. 3.2 Der ElliQ-Bot ist für ältere Menschen konzipiert. (Foto: ElliQ)

gelang, wendet Magic Leap sein Grundlagenwissen im Business-to-Business-Kontext an und stellte 2018 MICA eine ausgesprochen realitätsnahe Business-Assistentin vor, die nicht nur redegewandt, sondern auch attraktiv ist.

Eine derartige Körperlichkeit eines Assistenten, vermittelt einen gewissen Grad von Menschlichkeit. Das kann Sympathien wecken, führt aber auch zum sogenannten Uncanny-Valley-Effekt – einer kurzzeitigen Akzeptanzlücke. Je menschenähnlicher ein CUI visualisiert wird, desto eher erreicht man das unheimliche Tal. Anfänglich ist der Nutzer begeistert von der Menschlichkeit einer Maschine. Doch diese Begeisterung schlägt ab einem bestimmten Punkt ins Gegenteil um. Man erschrickt vor der menschlichen Nachbildung. Den Begriff des Uncanny Valleys und seinen Verlauf hat Masahiro Mori zuerst 1970 ausformuliert [3].

Gute Beispiele liefern die androiden Personen, die Hiroshi Ishiguro, Professor im Intelligent Robotics Laboratory an der Universität Osaka [4] entwickelt. Erica ist beispielsweise eine Ganzkörper-Figur mit dem Gesicht einer 23-jährigen Japanerin. Im Gespräch mit ihrem Gegenüber stimmen Mimik und Gestik mit einem echten Menschen überein. Doch wenn man ein Video mit dem humanoiden Roboter anschaut, setzt nach einigen Minuten ein Abstoßeffekt ein.

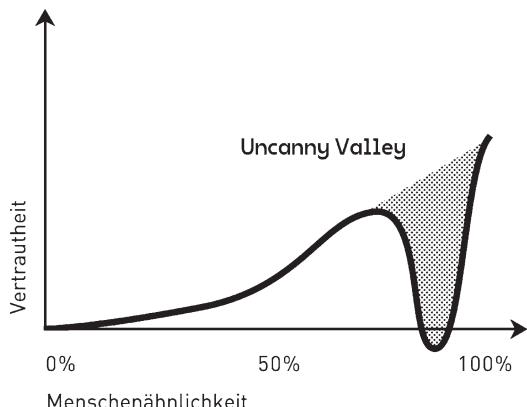
Menschenähnlichkeit

Das Uncanny Valley (siehe Abb. 3.3) kann man vermutlich nicht komplett umgehen, doch kann man es zeitlich nach hinten schieben. Je weniger ein CUI einem Menschen ähnelt oder menschliche Züge trägt, desto länger lässt das Uncanny Valley auf sich warten. Darauf setzt beispielsweise der chinesische Autohersteller Nio bei seinem CUI Nomi.

Dieser Assistent steckt zwar in einer beweglichen Kugel, die einem Kopf ähnelt, doch auf dem Display sind lediglich Augen zu sehen. Auf einen Mund oder die Nase haben die Designer bewusst verzichtet.

Damit stellt sich die Frage, warum CUIs überhaupt auf menschliche Äußerlichkeiten setzen sollten? Schließlich funktionieren sie mit Sprache. Ohne ein Gesicht sowie Körper

Abb. 3.3 Das Uncanny Valley



umgeht man das Uncanny Valley. Doch dabei verspielt man Sympathiepunkte. Zu sehen, mit wem man redet – überhaupt zu wissen, wo man beim Gespräch hinschauen muss – senkt die Hemmschwellen. Zudem ist man bei menschlicher Kommunikation gewohnt, dass der oder die Gegenüber mit dem Kopf nickt, lächelt oder sonstige nonverbale Signale sendet, dass sie noch immer zuhört und bei der Sache ist. Somit kann die menschliche Nachbildung bei einem CUI durchaus positive Effekte aufweisen.

Natürlich gibt es auch dafür Gegenbeispiele. Als Google im Mai 2018 seinen digitalen Assistenten Duplex vorstellte, waren etliche Berichterstatter erschrocken [5]. Man kann durchaus davon sprechen, dass sie das Uncanny Valley durchschritten. Der Sprachassistent rief während der Produktpräsentation bei einem Friseur in Kalifornien an und vereinbarte mit der Mitarbeiterin einen Termin. Dabei schob der digitale Assistent ein „Mmmmh“ ein, um zu verdeutlichen, dass sie wartet während die Mitarbeiterin im Kalender nachschaut. Google sagt dazu: „Wir entwerfen diese Funktion mit einer integrierten Offenlegung. So stellen wir sicher, dass sich das System ordnungsgemäß identifiziert. Was wir auf der I/O gezeigt haben, war eine frühe Technologie-Demo, und wir freuen uns darauf, Feedback in der weiteren Entwicklung des Produkts zu berücksichtigen“ [6].

Wie wir bereits zum Abschluss des ersten Kapitels darlegten, stellt Google Duplex sicherlich einen Meilenstein in der automatisierten Kommunikation dar. Das gilt auch für die Frage, wie „intelligent“ ein Assistent sein kann und sollte.

3.2.2 Wie intelligent soll Ihr Assistent werden?

Nach der Frage, wie der Assistent heißt und aussieht, ist die Frage nach der Intelligenz zunächst eine technische Frage. Werden ein Natural Language Understanding und ein Conversational Management eingebunden? Kann der Assistent also natürlichsprachige Äußerungen des Nutzers angemessen interpretieren? Kann der Assistent nutzwertige Antworten liefern und Handlungen auslösen? Noch einen Schritt weiter geht die Prediction, also das Treffen von Annahmen. Bucht ein Nutzer immer Bahnfahrkarten in der 2. Klasse mit einem Fensterplatz, könnte der Assistent dies bei der nächsten Anfrage als Annahme nutzen und fragen: „Wie immer in der 2. Klasse am Fenster?“

Noch weiter geht eine Sentiment-Analyse. Auf Deutsch könnte man Stimmungsdeutung sagen. Der Assistent kann aus der Modulation der Stimme Rückschlüsse auf die Stimmungen und Launen des Nutzers ziehen. Ein weiterer Aspekt der Intelligenz ist die Auswertung von Bildern. Das hängt natürlich in erster Linie mit den eingesetzten Geräten zusammen. Verfügt das Eingabegerät über eine Kamera? In dem Fall könnte eine Bilderkennung Objekte vor der Kamera deuten und vergleichbare Dinge suchen. Beim Thema Shopping ist das eine praktische Funktion. Wenn die Kamera aber das Gesicht des Nutzers erfasst, könnte der Assistent auch aus Mimik und möglicherweise aus der Gestik des Nutzers Rückschlüsse ziehen und diese in den Dialog sowie die Bearbeitung der Nutzeranfragen einbeziehen.

3.2.3 Unterhaltung und Funktionalität

Wo kann die Konversation stattfinden? Ist der Assistent besonders hilfreich, wenn man unterwegs ist, beispielsweise im Auto? Wird er auf dem Computer verwendet oder doch eher mit Smartspeakern oder Smart-TVs in den eigenen vier Wänden?

Da es keine schriftliche Anleitung gibt, muss das Onboarding des Nutzers möglichst einfach und reibungslos verlaufen. Auf dem Bildschirm eines Smartphones kann man noch Hilfestellung geben, aber bei einem Smartspeaker muss die Anfrage ohne derartige visuelle Hinweise funktionieren. Was kann der Nutzer fragen? Muss er oder sie bestimmte Formulierungen verwenden? An dieser Stelle muss der Assistent auch ein geeignetes Erwartungsmanagement betreiben. Er muss dem Nutzer klar und deutlich sagen, wobei er hilft und wobei er überfordert ist. Dient der Assistent dem Fahrkartenvverkauf im öffentlichen Nahverkehr, kann er keine Witze erzählen oder die Wettervorhersage liefern. An dieser Stelle sollte definiert werden, wie der Assistent mit Fehlern umgeht. Wie oft fragt er nach, bis er dem Nutzer sagt, dass er ihn nicht versteht bzw. nicht helfen kann?

Ein weiteres Stichwort in dieser Rubrik lautet Discoverability. Es beschreibt, wie Nutzer neue Funktionen bzw. neue Fragemöglichkeiten entdecken. Es geht nach einem ersten, zaghaften Ausprobieren darum, auch ein drittes oder viertes Mal den Assistenten zu benutzen. Er oder sie soll Teil des Nutzeralltags werden. Das funktioniert nur, wenn der Nutzer erinnert bzw. zum Ausprobieren neuer Funktionen und Anfragen motiviert wird. Daran schließt sich der Bereich Hilfe & Support an. Falls der Assistent mal nicht versteht, wo finden Nutzer Hilfe bzw. an wen können sie sich mit einem Feedback wenden?

3.2.4 Interaktion und Dialog

Die Stimme des Assistenten ist ein wesentlicher Punkt. Kann der Nutzer zwischen unterschiedlichen Stimmen – Mann/Frau – wählen? Nutzt man eine Standard-Stimme, die von einer Maschine generiert wird oder lässt man einen professionellen Sprecher oder Schauspieler alles aufnehmen? Über das Verfahren der Stimm-Synthese können virtuelle Stimmen aus echten Stimmen erzeugt werden und ein Sprachassistent kann wie ein bekannter Schauspieler klingen. Der Anbieter kann ganz bewusst eine neue Stimme schaffen, die seinen Anforderungen und Branding-Überlegungen entspricht, die quasi zu seiner Marke passt. Ohne Bildschirm und Körper verleiht die Stimme dem Assistenten im wörtlichen Sinne ein Gesicht. Man hat sofort eine Person vor dem geistigen Auge, wenn man eine Stimme hört. Mit der Stimmauswahl bestimmen Sie auch den Charakter des Assistenten. Es ist also eine Frage des Sonic Brandings.

Auch wenn bei CUIs gesprochene Sprache im Vordergrund steht, darf es visuelle Hilfe bei den Antworten geben. Insbesondere wenn es mehrere Antworten gibt oder eine Auswahl zu treffen ist, helfen sogenannte Canned Responses dem Nutzer. Er muss

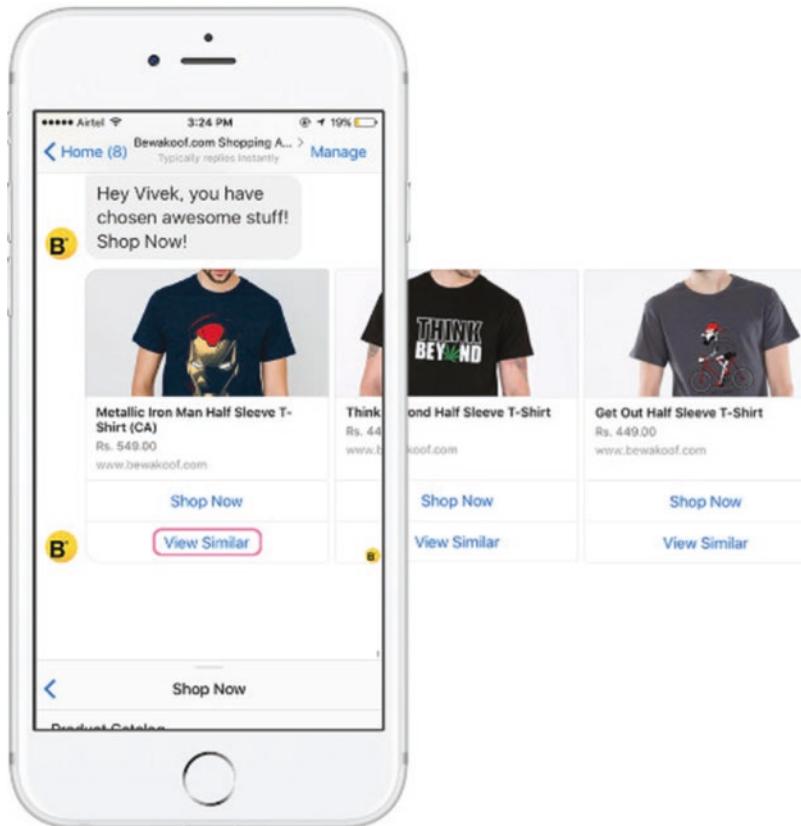


Abb. 3.4 E-Commerce im Messenger-Bot mit Auswahlmöglichkeiten (Bild: Bewakoof.com)

sich nicht die Optionen A, B und C merken, sondern kann auf dem Bildschirm vorformulierte Antworten nachlesen und dann auswählen. Das können auch Buttons, Karussells oder Texte sein, die man eigentlich aus der Welt der grafischen Benutzeroberflächen kennt. Einige Facebook-Messenger-Bots nutzen bereits Canned Responses (siehe Abb. 3.4).

Dabei stellt sich die Frage, wie der Assistent mit den Nutzern spricht. Wenn Nutzer Emojis nutzen, versteht der Assistent deren Bedeutung? Darf oder sollte er sogar mit Emojis antworten? Wie sieht es mit Fachbegriffen (ROI) und Kürzeln aus (MUC für München bei Flügen)? Duzt er den Nutzer oder wird dieser höflich mit Sie angesprochen?

Wie setzt sich der Assistent mit dem Nutzer in Verbindung, wenn er längere Zeit nicht genutzt wurde? Per Push-Benachrichtigung oder per E-Mail? Die Mail ist ja bereits ein Medienbruch, der sich aber im Sinne der Nutzungsquoten lohnen kann.

Analysewerkzeuge ermöglichen eine Auswertung der Nutzung und das Auslösen von Aktionen. Sinkt beispielsweise der Anteil der Nutzer auf mobilen Geräten bei einem Web-Shop, kann der Assistent Erinnerungen auslösen.

3.2.5 Monetarisierung und Bezahlung

Soll der Assistent etwas verkaufen oder dabei unterstützen? Dann sollte bereits in der Konzeptphase überlegt werden, wie und mit welchen Bezahlmethoden sich der Nutzer autorisieren kann. Wie funktioniert die Übergabe zum Bezahlsystem, dem Web-Shop und zurück zum Assistenten? Je einfacher das funktioniert und je niedriger die Hürden für den Nutzer ausfallen, umso eher kann ein Assistent eine große Hilfe bei E-Commerce-Anwendungen sein.

Viele der Elemente, aus denen Conversational User Interfaces bestehen, und die auf dem langen Weg von der gesprochenen oder geschriebenen User-Eingabe in natürlicher Sprache bis zum Ergebnis einer passenden Antwort oder einer ausgelösten Prozesskette, haben wir in diesem Kapitel beschrieben. Das Herzstück, die Verarbeitung von Sprache im Sinne eines rudimentären Verständnisses als auch der möglichst automatisierten Erzeugung von Antworten in natürlicher Sprache, fasst man unter dem Oberbegriff Natural Language Processing zusammen. Um dieses Feld geht es im nächsten Kapitel.

3.2.6 Monitoring und Reporting

Natürlich besteht ein Conversational User Interface auch aus vielen Elementen, die vom Anwender nicht direkt wahrgenommen werden. Darunter alle Funktionen, die die Inputs der Anwender tracken, die Dialoge zwischen CUI und Nutzer monitoren und in möglichst aussagekräftige Reports fassen, die von den Betreibern dann ausgewertet werden. Besonders weit entwickelte Systeme nehmen derartige Auswertungen automatisiert vor und weisen selbstlernende Funktionen auf, die Probleme teilweise sogar automatisch verbessern.

Literatur

1. Microsoft Cortana Demo auf der Build 2019, YouTube, <https://youtu.be/jIbkbYRgzD4>
2. Elli Q, <https://elliq.com>
3. Schwarz, Rob: 10 Creepy Examples of the Uncanny Valley, (25.11.2013), <https://www.stranger-dimensions.com/2013/11/25/10-creepy-examples-uncanny-valley/>
4. Intelligent Robotics Laboratory, Osaka University, <http://eng.irl.sys.es.osaka-u.ac.jp/>
5. Kremp, Matthias: Google Duplex ist gruselig gut, (09.05.2018), <https://www.spiegel.de/netzwelt/web/google-duplex-auf-der-i-o-gruselig-gute-kuenstliche-intelligenz-a-1206938.html>
6. Nieva, Richard: Google says it's designing Duplex with 'disclosure built-in' (11.05.2018) <https://www.cnet.com/news/google-says-its-designing-duplex-with-disclosure-built-in/>



Was sind NLP und NLG und wie funktionieren sie?

4

Zusammenfassung

Dreh- und Angelpunkt der CUIs ist natürliche Sprache. Die Assistenten verstehen menschliche Sprachbefehle und können auf gleichem Weg Antworten liefern. Die beiden Technologien Natural Language Processing (NLP) und Natural Language Generation (NLG) kommen dabei zum Einsatz. Doch wie funktionieren diese Techniken? Die renommierte indische IT-Wissenschaftlerin Ragmadura Mathangi erläutert in ihrem Beitrag, wie NLP und NLG die Assistenten zum Reden bringen. Mathangi arbeitet als Head of Data Science bei [PhonePe.com](#), einem mobilen Bezahldienst mit Sitz in Bangalore. Sie zählt zu den Top-Ten Data Scientists in Indien und ihr Name steht auf über 20 Patenten.

Wikipedia schreibt: „NLP (Natural Language Processing) ist die programmatische Methode, mit dem Computer Eingaben aus der menschlichen Sprache übernehmen können. Das Fachgebiet, das sich auf die Interaktion zwischen menschlicher Sprache und Computern konzentriert, wird als Natural Language Processing, kurz NLP, bezeichnet. Es befindet sich an der Schnittstelle von Informatik, künstlicher Intelligenz und Computerlinguistik.“ [1]

NLP brachte einen Paradigmenwechsel. Ohne die Fähigkeiten von NLP sind wir gezwungen, syntaktische Eingaben von Befehlen und Wortfolgen zu tätigen. Mit NLP können wir Computer durch natürliche Sprache steuern. Heute arbeiten wir mit Anwendungen, die nicht nur die menschliche Sprache verstehen (Natural Language Understanding),

Beitrag von Ragmadura Mathangi

sondern auch selbstständig Text generieren (Natural Language Generation). Es gibt oft eine Verwechslung zwischen NLP (Natural Language Processing), NLU (Natural Language Understanding) und NLG (Natural Language Generation). Am besten könnte NLP unterteilt werden in natürliches Sprachverständnis und natürliche Sprachgenerierung. Das Verstehen der natürlichen Sprache analysiert eingegebene Texte in Bezug auf inhaltliche Themen, Stimmung, Zusammenfassung usw. Die Generierung natürlicher Sprache hingegen taucht mehr in die Generierung des Textes selbst ein. Einige der automatischen Vorschläge für Antworten in E-Mail- oder Messaging-Systemen basieren auf der Generierung natürlicher Sprachen. 2018 wurde das Buch „I the Road“ [2] vom amerikanischen Verlag Jean Boîte Éditions veröffentlicht, der für das Werk in Anspruch nahm, dass es das erste Fiction-Buch sei, das ausschließlich mit Hilfe der natürlichen Sprachgenerierung geschrieben wurde. Im Bereich der automatisierten Generierung von faktischen Texten, beispielsweise der automatisierten Erstellung von Sport- oder Börsenberichten, wird schon seit Anfang der 2000er-Jahre experimentiert und selbst renommierte Publikationen, wie beispielsweise die New York Times, veröffentlichen immer wieder Stücke, die ausschließlich von Computern erzeugt wurden und offenbaren dies häufig erst später ihren Lesern.

Wir werden uns in diesem Kapitel ausführlich mit NLU und NLG beschäftigen, wobei der Schwerpunkt auf intelligenten Assistenzsystemen liegt. Das Kapitel ist nach der folgenden Pyramide aufgebaut. Oben in der Pyramide befinden sich die Algorithmen des maschinellen Lernens. Machine-Learning-Algorithmen (ML) sind das Rückgrat, um alle Data-Mining-Probleme zu lösen. ML-Algorithmen versorgen die „Technikschicht“. Sprachverarbeitungstechniken leiten alle Algorithmen aus der ML-Schicht ab, haben aber spezifische Strategien oder Nuancen, die auf „Sprachen“ oder „linguistischen Mustern“ basieren. In dieser Schicht befindet sich der Kern der natürlichen Sprachverarbeitung – sowohl NLP als auch NLG (siehe Abb. 4.1). Diese Ebene wiederum versorgt die Anwendungsfälle wie Suche, maschinelle Übersetzungen, intelligente Assistenten etc. Die Anwendungsfälle bestimmen die horizontale oder vertikale Ausrichtung in jedem Unternehmen.

Die Verarbeitung von natürlicher Sprache in Computern ist bereits heute allgegenwärtig. Wir erleben NLP, wenn wir mit einer Suchmaschine suchen und dabei nicht nur ein einzelnes Suchwort, sondern eine ganze Such-Phrase eingeben. Wir profitieren von maschineller Sprachverarbeitung, wenn uns Internetportale bestimmte Inhalte passend zu anderen Artikeln vorschlagen. Auch im Rahmen von Produktvorschlägen, die wir bei E-Commerce-Anbietern auf dem Bildschirm oder in Newslettern erhalten, steckt NLP „unter der Haube“. Auch bei der Kontrolle der von Nutzern eingegebenen Kommentaren und Bewertungen durch die Betreiber von Internetangeboten kommt NLP zum Einsatz. Natürlich auch, wenn Eingaben ins Smartphone gesprochen werden, sei es beim Ausfüllen von Formularen oder beim Eingeben von Beiträgen auf Messengerplattformen. Das gilt auch, wenn wir mit Sprachassistenten wie Alexa sprechen.

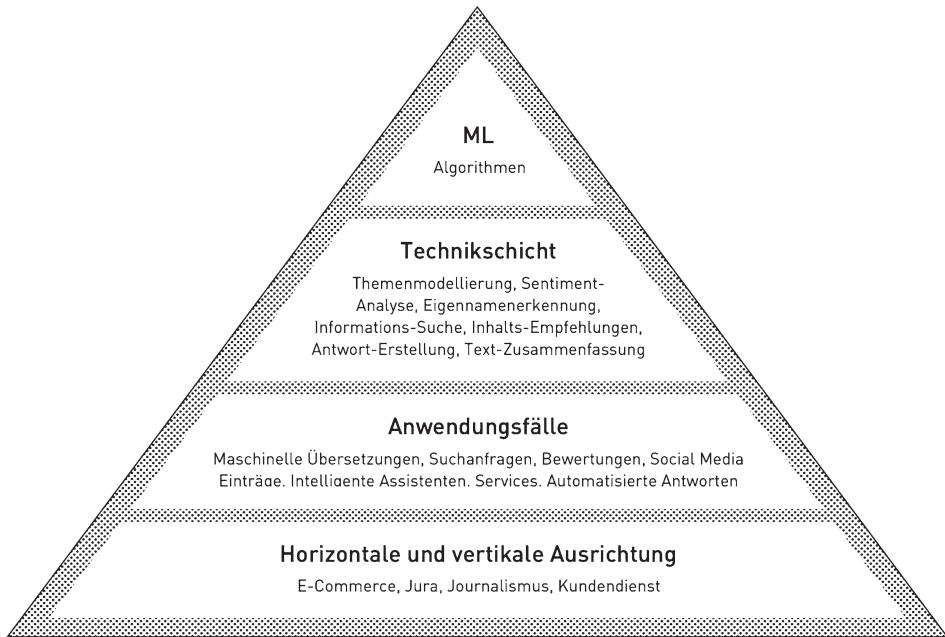


Abb. 4.1 In diesen Bereichen der Sprach-Pyramide wird NLU und NLG eingesetzt

Die beliebtesten NLP-Anwendungsfälle sind Suchen. Wir können die Geschichte der Suchmaschinen [3] bis Anfang 1945 zurückverfolgen, als ein amerikanischer Ingenieur, Vannevar Bush [4], einen Artikel veröffentlichte, um das in verschiedenen Aufzeichnungen verfügbare Wissen zu organisieren. In den 90er-Jahren gab es eine Reihe von Aktivitäten im Bereich der Suchmaschinen. Frühere Versionen von Suchmaschinen, die verwendet wurden, um Ergebnisse zu erhalten, indem sie nur „reguläre Ausdrücke“ auf die Dateinamen anwenden („The History of Search Engines – A Timeline“ [3]). Heutige Suchmaschinen verwenden fortschrittlichere maschinelle Lerntechniken, um die relevanten Ergebnisse zu erhalten und anzuzeigen. Suchmaschinen haben viele Dimensionen entwickelt, um nach Orten, Nachrichten, Bildern, Videos usw. zu suchen und um kontextuelle Ergebnisse zu liefern. Suchmaschinen kombinieren heute die Benutzermerkmale – wie Benutzereinstellungen und die Suchhistorie des Benutzers sowie die Benutzeranfrage, um die relevantesten Ergebnisse zu erhalten, so dass die Suchergebnisse für das gleiche Wort, beispielsweise „NLP“ vollkommen verschieden sein können, je nachdem, wann und wo es von unterschiedlichen Benutzern eingegeben wurde. Für einige Benutzer würden sich die Ergebnisse mehr auf die natürliche Sprachverarbeitung konzentrieren und für einige könnte es auf „Neurolinguistische Programmierung“ hinweisen. Heutige Suchmaschinen nehmen auch Spracheingaben entgegen und liefern dann die richtigen Ergebnisse. Es ist

bemerkenswert, dass 80 bis 85 Prozent der Anfragen in Google Zwei-Wort-Abfragen sind und mehr als Sechs-Wort-Abfragen machen nur 15 Prozent aus. Dies unterstreicht die natürliche Sprachkompetenz der Suchmaschinen [5].

Neben der Suche sind Maschinenübersetzungen derzeit ebenfalls bedeutende Felder von NLP: Wenn die Notwendigkeit die Mutter der Erfindung ist, dann ist die maschinelle Übersetzung die Mutter von NLP. Zunächst wie bei anderen natürlichen Sprachproblemen wurden die Sprachübersetzungsprobleme mit „Grammatikregeln“ gelöst. Diese Methode der Durchführung von Übersetzungen wurde äußerst komplex, als man beispielsweise versuchte, Sprachen wie Japanisch zu übersetzen. IBM begann in den 1990er-Jahren mit der Lösung maschineller Übersetzungen mit statistischen Methoden. Statistische maschinelle Übersetzung (SMT), wie sie genannt wird, hing nicht von linguistischen Regeln ab, sondern von Mustern, die von den Daten selbst dargestellt werden. Neuronale Maschinenübersetzer sind die neuesten Übersetzungsmechanismen, bei denen relativ tiefe neuronale Netzwerkmodelle verwendet werden, um Sätze von einer Sprache in eine andere zu übersetzen. Maschinelle Übersetzungen sind heute in professionellen Umfeldern nicht mehr wegzudenken. Die meisten international tätigen Unternehmen und Institutionen nutzen diese, um Inhalte für Mitarbeiter und Kunden weltweit zu lokalisieren. Aber selbst aus vielen privaten Situationen sind maschinelle Übersetzer heute ganz selbstverständlich im Einsatz. Längst haben sich Touristen daran gewöhnt, auf dem Smartphone nicht nur einzelne Worte, sondern ganze Sätze in Echtzeit zu übersetzen, und manche Situation im Restaurant oder Straßenverkehr konnte auf diese Weise wie von Zauberhand gelöst werden. Auch virtuelle Assistenten bieten Kundenservice in der Landessprache durch maschinelle Übersetzungen. Neben den großen amerikanischen Technologieunternehmen gibt es Alibaba, Baidu, PangeaMT, die zum Bereich der maschinellen Übersetzung beigetragen haben. Auch ein deutsches Unternehmen aus Köln bietet mit [DeepL.com](#) einen weltweit anerkannten Online-Übersetzer an. Maschinelle Übersetzung liefert auch den Nachweis dafür, dass sich NLP in den vergangenen Jahren erheblich verbessert hat. Vergleicht man die damals doch sehr mangelhaften Ergebnisse von Google Translate von vor wenigen Jahren und heute erkennt man die qualitativen Verbesserungen.

Die Erzeugung natürlicher Sprache unterscheidet sich vom Verständnis und der Analyse natürlicher Sprache, das wir bis hierher behandelt haben. Ein klassischer Anwendungsfall von Natural Language Generation (NLG) ist die Bereitstellung von Texten für automatisiert erzeugte Dashboards oder Berichte. Sogenannte automatisierte Insights: Wordsmith heißt beispielsweise das NLG-Softwareangebot des amerikanischen Unternehmens Automated Insights. In Abb. 4.2 ein Beispiel von der Unternehmens-Website, das zeigt, wie aus Datenpunkten das Verkaufs-Kommissionsmodell eines Unternehmens Diagramme erzeugt werden, die in der rechten Randspalte automatisiert erläutert werden und damit dem Betrachter einen besseren Einblick in die wichtigsten Aspekte der Datensammlung bietet (siehe Abb. 4.2).

Automatisiert erzeugte Rezensionen sind ein weiteres weit verbreitetes Anwendungsbereich von Natural Language Generation: 93 Prozent der online Kaufentscheidungen werden durch Reviews getroffen [6]. Reviews zu erzeugen und deren sinnvolle Präsentation

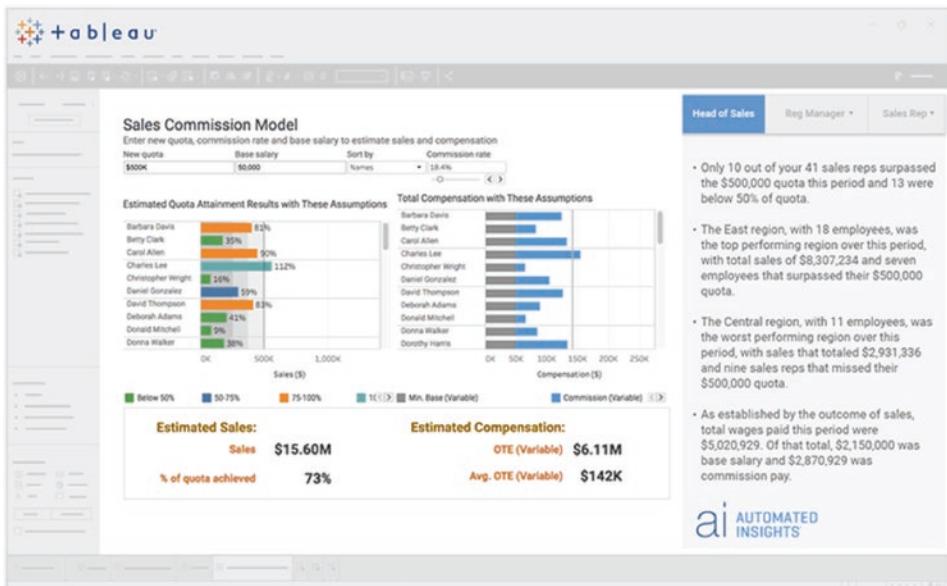


Abb. 4.2 Texterläuterungen durch NLG in einem Kommissionsmodell (Foto: Automated Insights)

auf den Websites von Reiseanbietern oder anderen E-Commerce-Anbietern haben einen großen Einfluss auf deren Umsätze. So verbessern beispielsweise die automatisierte Zusammenfassung und Hervorhebung der wichtigsten Kundenbewertungen und die Gruppierung von Bewertungen in negative und positive Beispiele das Kundenerlebnis und damit den Kaufprozess erheblich.

Auch im Feld von Social-Media-Monitoring findet Natural Language Processing Anwendung: Das Mining von Tweets und Facebook-Posts und die automatische Eskalation zu Servicemitarbeitern auf der Grundlage der darin erkannten Stimmung ist eine weitere weit verbreitete Anwendung von NLP (siehe Abb. 4.3). Mehr als 55 Prozent der Kunden wenden sich an Social-Media-Kanäle, um sich zu beschweren [7]. Stimmungs-Monitoring für Marken, das schnelle Erkennen von Beiträgen von Influencern, die über schlechte Erfahrungen mit der Marke berichten, Fragen der Markenpositionierung und alle geo-bezogenen Themen sind einige der Erkenntnisse, die Social-Media-Monitoring-Tools [8] unter Einsatz von NLP-Verfahrensweise liefern. Social Media könnte beispielsweise auch die Performance des Kundenserviceteams analysieren und damit Frühindikatoren erkennen.

Sowohl NLP und NLU als auch NLG bilden den Kern der intelligenten Assistenten. Diese müssen die Nutzerfrage verstehen, die Fragen einem Workflow-Prozess zuordnen und angemessen beantworten. Einige der intelligenten Assistenten bieten vordefinierte Antworten. In einigen Agenten werden die Antworten automatisch von der Maschine generiert und machen daher einen ziemlich realistischen Eindruck, so als hätte ein echter Mensch sie eingegeben. Siri, Googles Assistant und Amazon Alexa verwenden eine

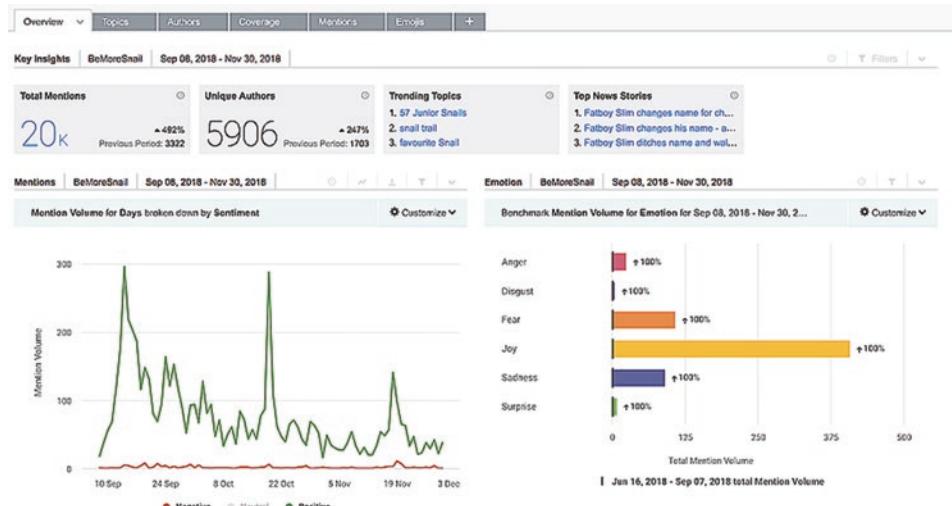


Abb. 4.3 NLP-Einsatz beim Social Media Monitoring (Foto: Brandwatch Consumer Research)

Kombination aus NLU und NLG. Siri und Google Assistant behaupten, dass sie auf jeweils einer halben Milliarde Geräten installiert sind. Auch die Zahlen von Amazon sind beeindruckend hoch. Die Genauigkeit, in der diese Systeme natürlichsprachige Nutzereingaben korrekt erkennen, hat sich innerhalb kurzer Zeit kontinuierlich und erheblich verbessert.

4.1 NLP-Anwendungen nach Branchen

Sogenanntes „Legal Tech“, also der Einsatz moderner Technologien im Bereich von juristischen Arbeitsprozessen, wird bis 2026 auf 3,7 Milliarden US-Dollar geschätzt [9]. NLP hilft Anwälten und Verbrauchern aktiv. Es gibt „Rechtsassistenten“, die in wenigen Sekunden Millionen von Dokumenten durchsuchen, die für die Recherche von juristischen Fragen benötigt werden. Nicht nur die Automatisierung der Forschung hat die Kosten gesenkt, sondern auch die Präzision, ähnliche Fälle wie den vorliegenden Fall zu finden. Die Vertragsanalyse unterstützt bereits heute Anwaltsteams und -organisationen auf der ganzen Welt bei der Identifizierung potenzieller Risiken und wichtiger Klauseln, die in einem Vertrag enthalten sind. NLP-Modelle, die die Vertragsanalysesoftware unterstützen, scannen Millionen von Verträgen und extrahieren die wichtigsten Merkmale des Dokuments. Derartige Software gibt die Daten in tabellarischer Form zur weiteren Analyse durch Anwälte aus Fleisch und Blut aus. Unternehmen wie Lex Machina prognostizieren das Ergebnis von Rechtsstreitigkeiten mit NLP und KI und helfen so Anwälten bei der Strategieentwicklung der Rechtsfälle.

Konkurrenzpreisvergleiche sind ein häufiger Anwendungsfall von NLP im E-Commerce. Ein E-Commerce-Unternehmen möchte die Preise seines Produkts mit denen eines Wettbewerbers vergleichen. In diesem Fall muss das E-Commerce-Unternehmen einen in einer Wettbewerber-Website erwähnten Artikel automatisch klassifizieren und mit seiner eigenen Website abgleichen. Amazon, Walmart, Target etc. scannen Tausende von Seiten anderer E-Commerce-Sites und passen ihren Produktpreis dem Markt über Millionen von Produktkategorien [10] in Echtzeit an. Die automatisierte Textgenerierung wird auch zur Erstellung von Produktbeschreibungen eingesetzt, z. B. für E-Commerce-Sites, bei denen es nicht nur um den richtigen Inhalt, sondern auch um die suchmaschinenoptimierte Verwendung von Keywords geht. Da die Regeln für Suchmaschinenoptimierung ziemlich komplex sein können, zeigt sich, dass Maschinen derartige Texte besser erstellen können, als es ein Autor je könnte. Auch hier gibt es international viele große und kleine Anbieter, darunter auch wiederum sehr innovative Unternehmen aus Deutschland (z. B. www.ax-semantics.com).

Die automatische Erstellung von Texten findet natürlich auch im bereits erwähnten Zusammenhang von Legal Tech eine Rolle, wo Mahnschreiben, Anspruchsschreiben, Widerspuchsmeldungen und ähnliches bereits heute in vielen Standardfällen, zum Beispiel beim Geltendmachen von Ansprüchen gegenüber von Fluggesellschaften bei Verspätungen und Flugausfällen, von spezialisierten, großen Anwaltskanzleien mittels NLG-Maschinen erstellt werden.

NLP findet auch im Journalismus Anwendung: Natural Language Understanding und Natural Language Generation wird verwendet, um große Mengen von Nachrichtenartikeln zusammenzufassen. Inshorts.com ist ein indisches Unternehmen, das Nachrichtenartikel für Benutzer in weniger als 60 Wörtern zusammenfasst, was insbesondere für die in Asien vorwiegende Nutzung von Smartphones mit kleinen Bildschirmgrößen ein attraktives Angebot ist. Resoomer.com ist ein französisches Unternehmen, das schnelle Textzusammenfassungen für jede Art von Text liefert. In den letzten Monaten haben Programme mehr als hundert Artikel in der renommierten Washington Post und anderen Tageszeitungen geschrieben [11]. Dies findet derzeit insbesondere in Themenbereichen wie Sport- und Wirtschaftsnachrichten statt, da in diesen Fällen eine brauchbare Geschichte durch Darstellung der Fakten (z. B. Tore, Torschützen und andere Spielgeschehnisse) erstellt werden kann. Die Themen erweitern sich jedoch stetig und umfassen Polizeiberichte (Mord, Waffe, Tatort, Umstände...), die ähnlich aufgebaut sind. Ergänzt werden können diese Beispiele von Robojournalismus um Tobi [12], ein textgenerierender Bot, der vom Zürcher Tagesanzeiger programmiert wurde und der bereits über 40.000 personalisierte Artikel über die Wahlen in der Schweiz „geschrieben“ hat.

Der Kundenservice eines mittleren oder größeren Unternehmens generiert Petabyte an Sprach- und Textdaten – in Form von Sprachanrufen, E-Mails, Tickets, Kundensupportchats etc. Um die Qualität der Kunden-Kontaktcenter zu gewährleisten, müssen Unternehmen ihre Kundenbetreuer regelmäßig beobachten und überwachen. Sie wollen die im Dialog mit Kunden anfallenden Daten auch verwenden, um zu verstehen, was ihre Kunden mögen und was nicht. Kundendienstdaten sind der Schlüssel zum Verständnis des NPS

(Net Promoter Score) eines Unternehmens. Der Net Promoter Score misst, wie wahrscheinlich es ist, dass ein Benutzer ein Produkt seinen Freunden und Verwandten empfiehlt. Text Mining der Kundensupportdaten hilft verschiedenen Unternehmen, die wichtigsten Treiber von NPS zu verstehen.

4.2 Wie funktioniert Natural Language Understanding?

Obwohl das Interesse und die dadurch getriebene Forschung an Natural Language Processing bereits bis in die 1950er-Jahren zurückreicht, sehen wir die weit verbreitete Nutzung von NLP im Alltag erst seit ungefähr 2012. Lange Zeit war das natürliche Sprachverständnis von Maschinen durch starre linguistische Regeln geprägt. Mit den riesigen Datenmengen allerdings, die seit Anfang der 2010er-Jahre weltweit gesammelt werden, spielen die Fortschritte in der Rechenleistung moderner Computer und die Fortschritte auf Seite der Programme und Algorithmen eine immer größere Rolle in der natürlichen Sprachverarbeitung. Big Data – also das Phänomen, dass wir immer weitere Teile unserer Kommunikation digital erledigen und auf diese Weise naturgemäß auch immer größere Anteile des Wissens in Form von Artikeln, Büchern, Verträgen und sonstigen Dokumenten in digitaler Form vorliegen, automatisiert untersucht und ausgewertet werden können, ist die Basis für die Algorithmen im Bereich der automatisierten Sprachverarbeitung. Das Ausmaß dieser Datenflut lässt sich nur schwer ermessen. Das World Economic Forum [13] veröffentlicht seit einiger Zeit dazu Zahlen. Demnach wurden im Jahr 2019 jede Minute 188 Millionen E-Mails verschickt und im gleichen Zeitraum 18 Millionen Textnachrichten und Neunzigtausend Tweets. Mit der wachsenden Datenmenge, die zur Auswertung zur Verfügung steht, wächst auch das Interesse von Forschern weltweit, anhand dieser Daten das Verständnis der Grundstrukturen von Sprache zu verbessern und so in Algorithmen verpacken zu können, damit Computer Sprache tatsächlich „verstehen“ können.

NLU-Techniken umfassen hauptsächlich Themenmodellierung, Extraktion von Wortteilen, Extraktion von benannten Entitäten (Informationsobjekten) oder Sentiment- und Meinungsanalyse. Die Informationsbeschaffung ist eine weitere Art des natürlichen Sprachverständnisses, bei der eine Anfrage oder ein damit zusammenhängendes Suchbegriffsdocument einbezogen wird.

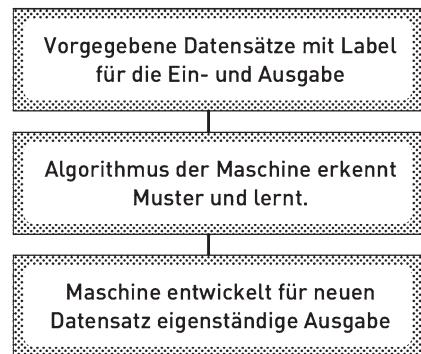
4.3 Was ist Machine Learning?

Bevor wir in die Nuancen der NLU-Techniken eintauchen, ist es wichtig, Machine Learning zu verstehen. Die heutigen NLP-Aufgaben werden weitgehend durch maschinelles Lernen angetrieben. Normalerweise programmiert man bei jeder Software-Anwendungsentwicklung den Computer, eine konkrete Aufgabe auszuführen. Beim maschinellen Lernen stellt man den Computern Daten in Form eines kontinuierlichen Stroms oder in Form von großen historischen Datenpools zur Verfügung. Man instruiert die Maschine, darin

enthaltene Muster zu identifizieren und ein Ergebnis aus diesen Mustern zu ziehen, um so selbst ein Programm zur künftigen Analyse von ähnlichen Datensätzen zu erstellen. Es gibt im Wesentlichen zwei Arten von Algorithmen, die Maschinen helfen, Muster aufzudecken: überwachte (supervised) und unbeaufsichtigte (unsupervised) Algorithmen.

Beim überwachten maschinellen Lernen wird der Maschine ein Datensatz sogenannter Labels zur Verfügung gestellt (siehe Abb. 4.4). Im Grunde genommen handelt es sich dabei um das erwünschte Ergebnis eines Algorithmus. Im Fall von NLU könnte es sich um einen zuvor von Menschen per Hand klassifizierten Text handeln. Die Maschine, die einen der Algorithmen verwendet, lernt das Muster in diesen Daten, um ein Ergebnis für einen neuen Datensatz vorherzusagen. Ein Beispiel: Wenn Leute „Mir geht es gut“ auf die Frage „Wie geht es dir?“ antworten, lernt die Maschine „Mir geht es gut“ als Ergebnis von „Wie geht es Dir?“. Wenn also das nächste Mal jemand fragt „Wie geht es dir?“, kann er mit „Mir geht es gut“ antworten. Ein weiteres Beispiel wäre die Vorhersage, ob es morgen Regenfälle geben wird oder nicht. Wenn wir der Maschine einen Datensatz an einem bestimmten Tag, einer Woche, einer bestimmten Zeit, der relativen Luftfeuchtigkeit, der Windgeschwindigkeit usw. zur Verfügung stellen, zusammen mit der Tatsache, dass es geregnet hat oder nicht, könnte die Maschine das Muster des historischen Niederschlagsverhaltens lernen und somit vorhersagen, ob es an einem bestimmten Tag regnen wird oder nicht. Große Systeme werden zunächst mit verfügbaren Daten trainiert („Kaltstart“). Diese Systeme werden jedoch intelligenter, sobald sie beginnen, echtes Feedback von den Kunden zu erhalten. Sie können sich mit den neuen Daten automatisch trainieren und werden intelligenter, wenn neue Daten eintreffen. Vor diesem Hintergrund ist es unvermeidlich, dass große Betreiber von Conversational User Interfaces Dialoge zwischen Nutzern und den NLP-Systemen von Mitarbeitern im Nachhinein abhören lassen und insbesondere in jenen Fällen, in denen der Dialog nicht abschließend erfolgreich war, die entsprechenden Passagen neu „labeln“, um dem Algorithmus die Möglichkeit zur eigenständigen Verbesserung zu geben. Amazon enthüllte beispielsweise, dass sie eine Stichprobe von Gesprächen aufzeichnete, um die Datenmodelle nachzutrainieren. Hier war die öffentliche Überraschung groß, obwohl ein derartiges Vorgehen beim Verfahren des Supervised Learning unumgänglich ist und, solange die abgehörten Passagen zuvor anonymisiert wurden, auch für die Nutzer unschädlich ist.

Beim unüberwachten maschinellen Lernen sind Labels für ein Ergebnis entweder nicht festgelegt oder der Maschine nicht bekannt. Der Algorithmus sucht im Datensatz systematisch nach einem zugrunde liegenden Thema oder einer Struktur, wenn ein Algorithmus Gruppen oder Segmente ähnlicher Kunden in einer großen Kundendatenbank finden soll, ohne zuvor zu wissen, nach welcher Art von Ähnlichkeit wir suchen – es also zunächst keine Ergebnisvariable gibt – wird ein unbeaufsichtigter Algorithmus verwendet (siehe Abb. 4.4). Die „Distanz“ zwischen den Kunden wird anhand ihres Verhaltens gemessen und ähnliche Kunden werden immer wieder zusammengefasst, bis sich eine klare Rangordnung der Ähnlichkeitsmerkmale ergeben. Übertragen auf Sprachbeispiele bedeutet dies, dass wir in einer Reihe von Dokumenten nach einer spezifischen Wortfolge suchen

Abb. 4.4 Unüberwachtes maschinelles Lernen

wollen, einem Satz, einem Ausdruck oder einem Synonym und nicht nur nach einem einzelnen Wort. Der Algorithmus untersucht alle Dokumente nach sämtlichen Wörtern, die im gesuchten Ausdruck enthalten sind und nach allen Wörtern, die dem betreffenden Wort ähnlich sind. Die Ähnlichkeit wird erneut gemessen als der „Abstand“ zwischen der Menge der Wörter im Suchbegriff und der Menge der Wörter in jedem Dokument.

4.4 Gängige NLP-Techniken

Bei der Sentimentanalyse oder dem Opinion Mining geht es darum, die Emotionen in einem Textkorpus zu analysieren, dabei gibt es viele Möglichkeiten wie eine Gefühlslage extrahiert wird. Der gebräuchlichste Weg ist die Verwendung eines lexikonbasierten Ansatzes. Hier werden Wörter mit den häufigsten positiven und negativen Wörtern verglichen und die Bewertungen werden auf der Grundlage der Treffer, die sie mit den positiven und negativen Lexika haben, bewertet. Raffiniertere Techniken versuchen, die Adjektive und Adverbien zu extrahieren und dann zu klassifizieren und so zu bewerten, ob ein Satz positiv oder negativ gemeint ist. Die Stimmungsanalyse könnte auch durch „überwachtes Lernen“ durchgeführt werden. Die Maschine wird auf eine Reihe von „positiven“ und „negativen“ Sätzen trainiert. Angesichts dieser positiven und negativen Beispiele kann die NLP-Engine alle positiven und negativen Sätze einordnen. Neben der bloßen Klassifizierung von Gefühlen als positive und negative Sätze versucht man in der Praxis sowohl andere Dimensionen von Emotionen (Wut, Ekel, Angst, Glück, Traurigkeit und Überraschung), als auch Subjektivität der Meinungen (von sehr subjektiv bis sehr objektiv) zu identifizieren. Zum Beispiel: „Die Kameraauflösung des Telefons mit zehn Megapixel ist die höchste auf dem Markt.“ ist eine eher objektive Aussage im Vergleich zu „Das Telefon hat schöne Funktionen, wie zum Beispiel eine gute Kamera“. Eine weitere verbreitete NLP-Technik ist die Named Entity Recognition („Erkennung benannter Objekte“). Nehmen wir als Beispiel eine Filmkritik aus der Internet Movie Database zur TV-Show „The Big Bang Theory“ [14]. Der Nutzer/Autor schreibt:

"Eine der besten TV Sitcoms.

18. Dezember 2007 | von hfan77 - Alle meine Bewertungen ansehen

Ich habe ein paar Episoden der neuen CBS-Sitcom **The Big Bang Theory** gesehen und für mich ist sie eine der am besten geschriebenen Sitcoms, die derzeit im Fernsehen läuft. Die Skripte sind gefüllt mit wissenschaftlichen Referenzen, die einige Leute vielleicht nicht verstehen, aber sie sind sehr witzig. **Johnny Galecki** und **Jim Parsons** sind definitiv eines der aktuellsten Comedy-Teams im Fernsehen, die über eine großartige Chemie verfügen und durch die Aufnahme von **Kaley Cuoco** als Nachbarin Penny die Show aufwerten. Nicht zu vergessen die Nebendarsteller mit **Simon Helberg** als Wolowitz, der ein wenig aussieht wie **Paul Reubens**, der Mann hinter **Pee-wee Herman** und **Kunal Nayyar** als ihr Freund aus Indien.

Wenn Sie sich den Abspann genau ansehen, werden Sie feststellen, dass es einen Hinweis gibt, der besagt: "Science Consultant". Es ist die erste Sitcom, die einen solchen Consultant einsetzt. Es war eine sehr gute Idee von **Chuck Lorre**, dem Schöpfer der Show, der auch hinter dem CBS-Hit **Two and a Half Men** steckt, einen zu engagieren, der wissenschaftliche Ehrlichkeit in die Show bringt.

Ich hoffe, dass es die **The Big Bang Theory** für mehrere Jahre gibt. Sie hat eine große Zukunft vor sich."

Algorithmen extrahieren in diesem Fall die Charaktere oder die Szenen über die der Rezensent spricht, zusammen mit dem Sentiment für die benannten Schauspieler und Serien. Diese Algorithmen analysieren die relative Wortposition jedes Wortes in einem Satz und treffen daraus abgeleitete Aussagen, welches „Wort“ eine Person, einen realen Ort, eine Organisation, eine Filmlocation, ein Datum usw. sein könnte. Es gibt mehrere Techniken, um eine derartige Erkennung benannter Entitäten zu lösen. Eine Methode heißt „Look-up based“. Dabei handelt es sich um eine recht einfache Möglichkeit, eine Liste von Orten, Organisationen und Personennamen zu erstellen und eine nachschlagende Suche durchzuführen. In unserem Fall könnte dies eine Liste von Charakteren und Darstellen der Show sein. Dieser Ansatz ist jedoch im Allgemeinen nicht sehr skalierbar und rechenintensiv. In unserem Fall handelt es sich ja um ein geschlossenes Themenuniversum und daher ist ein Lookup-basierter Ansatz vielleicht keine schlechte Idee. Eine derartige Darstellerliste würde eine endliche Anzahl bekannter Einträge aufweisen. Eine weitere Methode nutzt sogenannte vorgeschulte Klassifikatoren: Es gibt viele vortrainierte Klassifikatoren wie Stanford-NER, NLTK NER, SPACY usw., die öffentlich zugängliche Text-Korpusse zum Training ihrer Machine Learning Modelle verwendet haben. Den Maschinen wurde ein riesiger Korpus an Wikipedia-Einträgen, Nachrichten oder ähnlichen Daten zusammen mit daraus manuell extrahierten „Entitäten“ zugeführt. Diese Klassifikatoren, wenn sie ein neues Wort erhalten, können also klassifizieren, ob es sich um einen Ortsnamen, Personennamen usw. handelt. Vorgeschulte Klassifikatoren lösen nicht alle

Aufgaben. In vielen Fällen sind ihre Genauigkeiten auch nicht gut. Normalerweise wird bei einer NER-Extraktion ein kombinatorischer Ansatz verwendet. Zum Beispiel Nachschlagen plus vortrainierter Klassifikator.

Auch hier kommt das bereits beschriebene überwachte Lernen zum Einsatz. Ich habe den Text hervorgehoben, um zu zeigen, dass man einen NER trainieren kann, verschiedene Typen von Namen zu erkennen – diejenigen, die gelb markiert sind, sind Schauspielernamen, Charaktere in Orange, Organisationsname in Grün, der Showname in Rot und der Drehbuchautorennname in Blau. In diesem Fall könnten wir unseren eigenen benutzerdefinierten, überwachten Klassifikator bauen, indem wir der Maschine die Sätze zuführen, die mit den Charakternamen beschriftet sind und diejenigen, die es nicht sind. Hier würde dies bedeuten, dass wir händisch einen „Trainingskorpus“ erstellen müssten, bei dem jedes Wort in einer Rezension ein Charaktername, Schauspielername oder Organisationsname ist. Das ist ein sehr teurer und zeitaufwendiger Prozess und sollte erst dann eingesetzt werden, wenn alle anderen Methoden nicht funktionieren.

Neben der beschriebenen Sentimentanalyse und der Erkennung benannter Objekte, stellt die Informationsbeschaffung über Suchmaschinen den derzeit vermutlich größten Anwendungsfall von Natural Language Processing dar. Bei einer Suchabfrage besteht die Aufgabe darin, eine Reihe von Dokumenten in einer Trefferliste zurückzugeben, die für den Suchbegriff relevant sein könnten. Suchmaschinen führen drei Vorgänge durch: Crawlen, eine Art Durchforsten aller vorhandener Websites, Indizieren und Ranken von Treffern [15].

Bots durchsuchen ständig die Websites von einem Link zum anderen, stoßen dadurch auf neue Webseiten oder aktualisieren bestehende mit einer vorgegebenen Häufigkeit. Diesen fortlaufend durchgeführten Prozess nennt man Crawling

Suchmaschinen speichern und organisieren die während des Crawling-Prozesses gefundenen Inhalte. Dokumente werden so abgelegt, dass sie bei einer Suchanfrage schnell abgerufen werden können. Dokumente werden typischerweise rückwärts indiziert und der Abgleich eines Suchbegriffs mit einem Index erfolgt in extrem kurzer Zeit. Die Dokumente, die einem Index entsprechen, werden zurückgegeben. Ein Beispiel ist in Abb. 4.5 dargestellt [16].

Worte	Dokument
der	Dokument 1, Dokument 3, Dokument 4, Dokument 5, Dokument 7
Doktor	Dokument 2, Dokument 3, Dokument 4
sagt	Dokument 5

Abb. 4.5 Indexierung von Dokumenten

Der umgedrehte Index ist ein schneller Weg, um Dokumente für ein abgefragtes Wort abzurufen. Es gibt auch andere Formen der Indexierung. Aber das ist die am häufigsten verwendete Variante.

Die Suche nach dem Rückwärtsindex erfolgt mit einer Reihe von Algorithmen wie word2vec, KDTree etc. Ein Beispiel für einen Ranking-Algorithmus ist der „Page Rank“-Algorithmus, der das Ranking einer Seite in einem iterativen Prozess bestimmt, basierend auf der Rangfolge der Links, die auf die Seite kommen und der Anzahl der Links, die von der Seite zeigen. Ebenfalls berücksichtigt wird der „Page Rank“ der Seiten, die auf diese Seite verweisen. Für ein relevantes Ranking wird der Text oder der Index der Seite zusammen mit anderen Personalisierungsfaktoren wie Standort, historische Suche usw. bewertet.

Intent Mining ist ein weiterer bedeutender Anwendungsfall für Natural Language Understanding. Intents sind das, was der Kunde eines Unternehmens ansprechen oder lösen will. Der Intent ist der Grund, warum ein Kunde das Unternehmen oder die Organisation kontaktiert. Schreibt ein Kunde beispielsweise eine Beschwerde-E-Mail, wendet sich der Kunde an ein Unternehmen via Chat. Ruft ein Kunde in einem Callcenter an oder gibt ein Kunde eine Bewertung eines Produkts auf einer Website ab, möchten Unternehmen automatisiert verstehen, was sein Anliegen – der Intent – des Kunden ist, um basierend darauf geeignete Kommunikationskanäle und -strategien zu beschreiten. Intent Mining kann sowohl durch überwachte als auch durch unüberwachte Techniken durchgeführt werden. Im Falle von überwachten Techniken annotierte man in der Vergangenheit vorgebrachte Beschwerden und Kundenäußerungen mit einer bestehenden Ontologie. Sobald dieser Taggingvorgang abgeschlossen ist, wird ein Algorithmus eingesetzt, um die darin enthaltenen sprachlichen und inhaltlichen Muster zu analysieren und diese Muster in den aktuellen Kundenäußerungen aufzuspüren. Das Ergebnis von Intent Mining bei einem typischen E-Commerce-Anbieter sieht häufig wie in Abb. 4.6 aus.

Die häufigste Frage dreht sich um die Möglichkeit, erworbene Produkte zurück zu geben. Am zweithäufigsten fragen Nutzer nach Coupons. Um allerdings auf konkrete Nutzerbedürfnisse auch eine befriedigende automatisierte Antwort erstellen zu können, müssen die Fragen natürlich nicht nur statistisch, sondern in jedem Einzelfall genau erkannt werden. Der folgende Prozess beschreibt die Intent-Mining-Schritte in Chat/Sprach-Bots unter Verwendung von überwachtem Lernen. Wir haben gesehen, wie überwachtes Lernen in anderen Techniken wie Named Entity Recognition, Sentiment Mining etc. eingesetzt werden kann. Hier ist ein detaillierter, schrittweiser Prozess des überwachten Lernens notwendig.

Alle bislang behandelten Aspekte von Natural Language Understanding bezogen sich auf das Verarbeiten und Verstehen von Textdaten. Bei Callcentern und im Bereich von Smartspeakern dreht sich alles um gesprochene Sprache. Es gibt eine Menge Audio-Spracherkennungssoftware, die automatisch gesprochene Sprache in Text umwandelt. Der wird dann mithilfe der beschriebenen Natural-Language-Processing-Methoden ausgewertet.

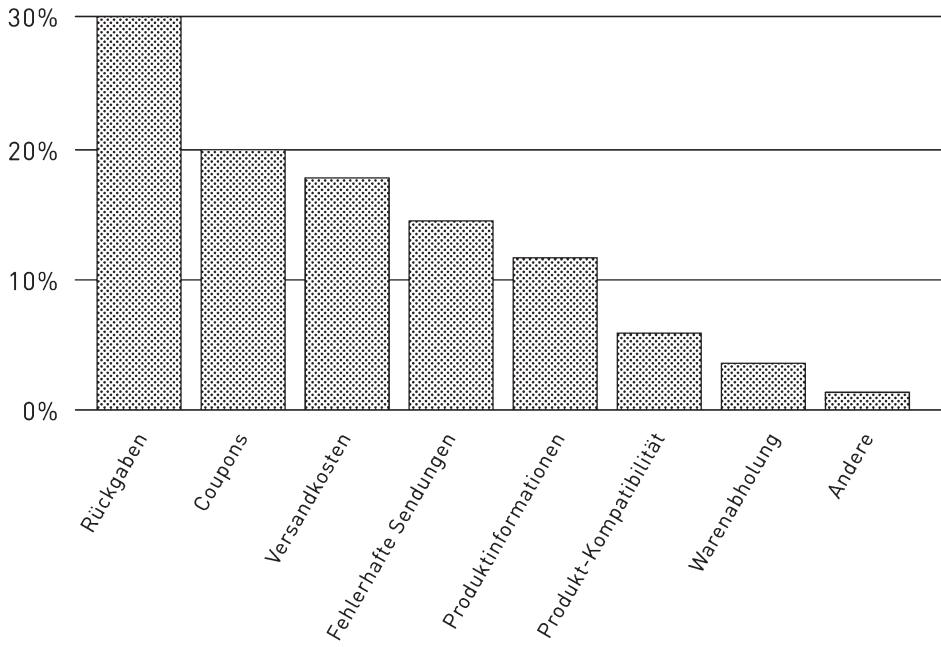


Abb. 4.6 Typische Verteilung der Gründe für Kundenanfragen an ein Unternehmen

4.5 Wie funktioniert Voice to Text?

Die Übertragung von gesprochener Sprache in Text nennt man Automatic Speech Recognition (ASR). Die kleinste bedeutungsunterscheidende sprachliche Einheit bezeichnet man als Phonem (z. B. *b* in „Bein“ im Unterschied zu *p* in „Pein“). Eine maschinell lernbasierte akustische Analyse wandelt Sprache in Phonemdaten um (siehe Abb. 4.7). Firmen wie Google, Amazon etc. haben einen riesigen Korpus an Audiodateien gesammelt und in Phoneme umgeschrieben. Mit diesen Lerndaten können Phonem-Sätze für neue Audiodaten transkribiert werden. Sobald eine Sprachaudiodatei in Phoneme umgewandelt wurde, wird das wahrscheinlichste Wort, das ein Phonem bilden konnte, erkannt. Beispielsweise könnte eine Ausgabe von ASR sein [17]:

- HHHEE_LL_LLLOOO
- HHHUU_LL_LLLOOO
- AAAUU_LL_LLLOOO

Nachdem wiederholte Buchstaben und Leerzeichen bereinigt wurden, sehen die Wörter wie Hello, Hullo und Aullo aus. Von diesen ist jedoch das mögliche Wort im Dictionary Hello. Daher wird das Wort „Hello“ gewählt. Sobald einzelne Wörter

Abb. 4.7 Sprache in Phonemdaten umwandeln

in einer Nutzeräußerung herausgearbeitet sind, wird ein richtiger Satz mit Hilfe von Sprachmodellen „korrigiert“. Die folgende Ausgabe von einem ASR könnte sein:

“How pan i help you”

“How van i help your”

Ein Sprachmodell betrachtet jedes Wort und die Wahrscheinlichkeit, dass ein Wort einem anderen Wort folgt. Aus einem bestehenden Datenkorpus errechnet es die Wahrscheinlichkeit, dass „I“ „pan“ folgt und andererseits die Wahrscheinlichkeit, dass „I“ auf „van“ folgt und sucht in der Liste der Worte nach Wörtern die mit deutlich höherer Wahrscheinlichkeit auf „I“ folgen. Basierend auf den Wahrscheinlichkeiten korrigiert es „pan“ und „van“ zum Wort „can“. Somit werden nach Anwendung der Sprachmodelle beide Sätze wie folgt korrigiert:

“How can I help you”

NLP ist das Herzstück der Dialogsysteme, die wir als Conversational User Interfaces und intelligente Assistenten kennen. Dialogsysteme unterstützen Natural Language Bots dabei, Fragen im Freistil- oder Open-Text-Format zu stellen. Dialogsysteme lassen sich grob in zielorientierte Systeme und Allzweck-Systeme einteilen.

Allzweck-Dialogsysteme: Das sind lustige und interaktive Bots. Sie arbeiten nicht mit hoher Genauigkeit. Diese Systeme versuchen, den Benutzer engagiert zu halten und der Benutzer hat möglicherweise kein bestimmtes Ziel vor Augen. Daher können Fragen zu allem möglichen auftreten und die Antworten können ebenfalls ziemlich vage sein. Diese Bots können auch verwendet werden, um Daten über Benutzergespräche zu sammeln.

Ein Beispiel

- Nutzer: *Was ist Dein Name?*
- CUI: Das habe ich Dir bereits gesagt, Hans.
- Nutzer: *Nein, ich meine Dein richtiger Name.*
- CUI: Stephanie ist mein richtiger Name.
- Nutzer: *Ich dachte er sei Hans?*
- CUI: Wie kommst Du denn darauf?
- Nutzer: *Weil Du es gesagt hast.*
- CUI: Ich muss meine Hausaufgaben machen.

Regelbasierte Systeme sind in der Regel die treibende Kraft solcher Q&A-Systeme. Es gibt Vorlagen für Fragen und Vorlagen für Antworten. Auch wenn diese Systeme häufig durch launige Template-Antworten versuchen, unterhaltsam zu erscheinen, sind diese aus verschiedenen Gründen derzeit nicht flexibel und auch nur begrenzt nützlich.

Neben diesen häufig nur begrenzt nützlichen allgemeinen Dialogsystemen gibt es die sogenannten zielorientierten Dialogsysteme: Dies sind virtuelle Assistenten für kommerzielle Zwecke oder intelligente Assistenten, bei denen die Benutzer ein Problem lösen wollen. Die Genauigkeit dieses Systems ist wichtig, will man die Nutzererwartungen auch nur ansatzweise befriedigen. Die Fragen stammen zwangsläufig aus einem bestimmten Bereich wie Flugreservierungen, Banken, Versicherungen, Dienstleistungen usw. Die Antworten müssen sehr spezifisch und zielgerichtet sein.

Beispiel

- Assistant: Guten Tag Frau Müller. Willkommen beim XY-Haarsalon. Was kann ich heute für Sie tun?
- Nutzer: Ich möchte die Spitzen kürzen und Strähnen färben.
- Assistant: An welchem Tag wollen sie zu uns kommen?
- Nutzer: Am Mittwoch kommender Woche.
- Assistant: Sie meinen am 10. Januar?
- Nutzer: Ja.
- Assistant: Um welche Uhrzeit wollen Sie zu uns kommen?
- Nutzer: Am Vormittag gegen 10:00
- Assistant: Das ist möglich. Ich habe also für Sie einen Termin zum Spalten kürzen und Strähnen färben für Mittwoch 10. Juli um 10:00 aufgenommen. Kann ich noch etwas für Sie tun?
- Nutzer: Nein.
- Assistant: Bitte geben Sie uns Bescheid, wenn Sie den Termin nicht wie vorgesehen wahrnehmen können.



Abb. 4.8 Workflow für Reservierung im Restaurant

Ein intelligenter Assistent beantwortet Freitext-Fragen vom Kunden. Der virtuelle Assistent prüft die Art des Kundenwunsches und bucht einen Termin nach Bestätigung von Datum und Uhrzeit, wobei der Fragensatz und die Antworten für jede Absicht als „Workflow“ bezeichnet werden.

In Abb. 4.8 ist die „Reservierung eines Restaurants“ beabsichtigt. Angenommen, der Kunde hätte gesagt: „Ich möchte heute Abend Chinesisch in der Innenstadt essen“, hätte der Bot dies als eine andere Absicht betrachtet und daher einen anderen Workflow angesteuert. Der Bot würde in diesem Fall direkt die Verfügbarkeit prüfen und eine entsprechende Antwort geben. Im Falle von Voice-Bots erledigt der Bot eine erste Aufgabe von Voice to Text und identifiziert dann eine Absicht, gefolgt von der Durchführung eines dafür vorbereiteten Workflows. Die Erkennung von Intents (Essen gehen = Restaurant-Reservierung) und Entities (Chinesisch, Innenstadt) wird bei diesen Systemen in der Regel durch den Einsatz von überwachten Modellen gelöst. Ein separates System verfolgt den Workflow und liefert entsprechende Antworten.

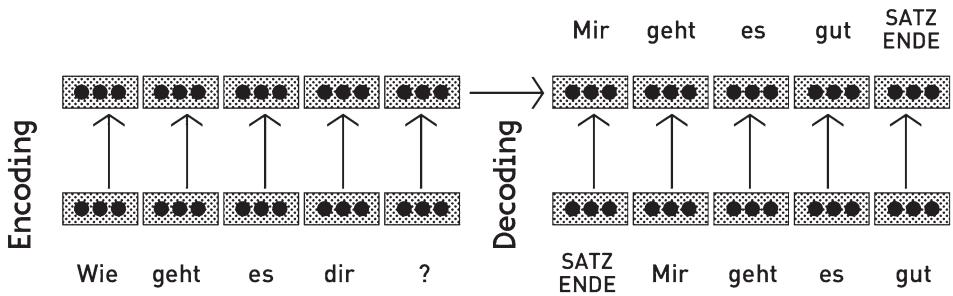


Abb. 4.9 Encoding und Decoding von Fragen

Ein weiterer Ansatz, nützliche Antworten in zielorientierten Systemen zu liefern, besteht darin, sehr große Textkorpusse zum Training heranzuziehen und einen Algorithmustyp einzusetzen, der ursprünglich aus dem verwandten Feld maschinelner Übersetzungen stammt: Die sogenannte Sequenz-Sequenz-Methode. Bei Sequenz-zu-Sequenz-Algorithmen sind die Ausgänge sowohl vom Eingang als auch von den vorherigen Ausgängen abhängig. Die in Abb. 4.9 gezeigten Ein- und Ausgabesätze können unterschiedlich lang sein. Bei einer Frage bestehend aus drei Worten kann es eine Antwort mit zwei Wörtern geben. In diesem Verfahren gibt es drei Komponenten – den Encoder, den codierten Ausgang (Vektor) und einen Decoder. Der Encoder verarbeitet die Eingangsdaten oder Fragen. Dann kodiert es den Satz mathematisch in einen Vektor und dieser Vektor wird in den Decoder eingespeist. Der Decoder gibt den Satz aus. Wie in jedem anderen maschinellen Lernsystem erfolgt das Training mit einer Vielzahl von Input- und Output-Antworten. Sobald die Architektur das Muster erlernt hat, ist sie in der Lage, angemessene Antworten auf unterschiedliche Eingabefragen zu geben.

Die bereits erwähnte Mehrdeutigkeit in der natürlichen Sprache legt es nahe, bei jedem Schritt im Dialog die Wahrscheinlichkeit für den aktuellen User-Wunsch (Intent) erneut zu überprüfen und diesen zu dokumentieren. In den letzten Jahren haben sog. Dialog State Tracking Challenges einen gemeinsamen Prüfstands- und Bewertungsrahmen für diese Aufgabe sowie große Mengen gelabelte Dialogdaten bereitgestellt, die zum Training der Systeme genutzt werden können.

Bereits am Anfang des Kapitels sprachen wir davon, dass auch Inhaltsempfehlungen, wie wir sie häufig bei Nachrichtenwebsites oder bei Produktempfehlungen von E-Shops kennen, ein bedeutendes Anwendungsgebiet von NLP darstellen: Empfehlungssysteme werden im Allgemeinen in drei Arten eingeteilt: artikelbasiert, nutzerbasiert und inhaltsbasiert. Die Stärke von NLP kommt bei inhaltsbasierten Empfehlungen zum Zuge. Ein Dokument, zum Beispiel ein Artikel, kann zu mehreren Themen gehören. Die NLP-Engine erkennt, dass ein bestimmter Nachrichtenartikel 30 Prozent Sport, 20 Prozent Wissenschaft, 10 Prozent Recht enthält. Millionen von Artikeln werden zunächst zu verschiedenen Themen zusammengefasst. Latent Dirichlet Allocation (LDA) ist eine gebräuchliche

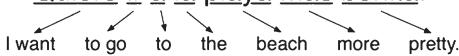
Technik, um Texte zu gruppieren. In LDA bilden Wortgruppen ein Thema. Eine Gruppe von Themen bildet ein Dokument (in unserem Fall den Nachrichten-Artikel). Diese Dokumente haben sogenannte Viele-zu-Viele-Beziehungen (n:m-Beziehung) zu Themen und Themen haben Viele-zu-Viele-Beziehungen zu Wörtern. Ein Artikel soll unterschiedliche Gewichte für verschiedene Themen haben, wie im obigen Beispiel gezeigt. Ähnliche Wörter könnten zu unterschiedlichen Themen gehören – zum Beispiel das Wort „Führung“ könnte zu einem Thema aus Politik, Wirtschaft, Wissenschaft oder Sport gehören. Die LDA ordnet Wörter den Themen und Themen den Dokumenten zu. Wenn ein Leser den Artikel beendet oder er ihm gefällt, werden Artikel, die diesen Artikel in Bezug auf die Verteilung der Themen ähnlich sind, aus der Datenbank geholt und dem Benutzer angezeigt.

4.6 Erzeugung von Texten – Natural Language Generation

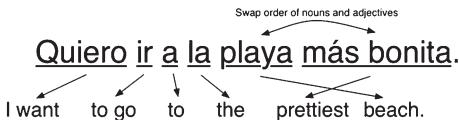
Während es im Feld des Natural Language Understanding darum geht, Texte und Nutzeräußerungen zu analysieren und richtig einzuordnen, geht es bei NLG um die Herausforderung der maschinellen Erstellung des Textes. Einige Beispiele sind die Erstellungen von Berichten oder textlichen Beschreibung aus tabellarischen Daten und anderen Abbildungen. Die Erstellung von Schlagzeilen für einen Nachrichtenartikel, die Zusammenfassung eines bestehenden Textes sowie maschinelle Übersetzungen. Wir werden die Techniken für jeden der Anwendungsfälle beschreiben.

1. **Automatisierte Insights:** Datenmengen werden zusammengeführt und es werden Mengen an Berichten für den gegebenen Datenpool erstellt – dies können beispielsweise Sportstatistiken aller Spieler und Spiele sein. Die Maschine sucht dann nach einem signifikanten Muster (Sharp increase oder Sharp decrease) und bereitet dann die Erkenntnisse nach vordefinierten Vorlagen vor. Vorlagen können vordefiniert werden oder Maschinen können sich die von Menschen geschriebenen Erkenntnisse ansehen und automatisch Vorlagen erstellen.
2. **Textzusammenfassung:** Dies wird entweder durch extraktive Textverdichtung oder abstrakte Textverdichtung erreicht. Bei der extractiven Methode werden Sätze automatisch gruppiert und der relevante Satz in jeder Gruppe als Zusammenfassung ausgewählt. Bei der abstrahierenden Methode müssen die als Zusammenfassung ausgewählten Sätze nicht Teil der ursprünglichen Botschaft sein, indem ein Korpus von Sätzen mit einer verfügbaren Zusammenfassung oder Überschrift trainiert wird. Deep Learning – speziell die Reihenfolge, in der die Modelle zu sequenzieren sind, wird verwendet, um die Modelle zu trainieren, und sobald das Training abgeschlossen ist, können wir dann für einen bestimmten Textabschnitt die Zusammenfassung automatisch erhalten. Bei dieser Methode können diese Sätze Teil des Originaltextes sein oder nicht.

3. Maschinenübersetzungen: Die einfachste Art der maschinellen Übersetzung ist die „Rules based machine translation“ (RBMT). Hier übersetzen wir zunächst Wort für Wort (ein Dictionary-Lookup) und wenden die Struktur der Eingabe- und der Ausgabesprache an. Als Beispiel betrachten Sie die Wort-für-Wort-Übersetzung von Englisch nach Spanisch. [17]

Quiero ir a la playa más bonita.


Der englische Satz ergibt keinen Sinn. Daher wird eine zusätzliche grammatischen Rekonstruktion angewendet und der Satz lautet nun: „Ich will zum schönsten Stand gehen.“

Quiero ir a la playa más bonita.


Dieser Prozess ist jedoch nicht skalierbar, da die Sätze komplexer werden, müssen auch die Übersetzungsregeln extrem vielseitig sein. Der nächste Meilenstein der maschinellen Übersetzung wurde mit statistischen Modellen erreicht. Hier wird die Wahrscheinlichkeit von Wörtern/Phrasen verwendet, um Wörter/Phrasen von einer Sprache in eine andere zu übersetzen. Statistische Maschinenübersetzungen ebneten den Weg für „Neuronale Maschinenübersetzungen“ (NMTs). NMTs sind hochmoderne maschinelle Übersetzungssysteme, die auf einem neuronalen Netzwerk basieren. Die Maschine wird mit Beispielen von Sätzen aus einer Sprache als Input trainiert, wobei der Output Sätze aus einer anderen Sprache sind. Eines der Probleme, die die Informatik bei Machine Translation gelöst hat, ist die variable Länge der Ausgabe. Betrachten wir zum Beispiel das Paar Englisch-Deutsch-Sätze.

We won. Wir haben gewonnen.

Die zwei englischen Wörter haben drei deutsche Wörter als Ausgang. In einigen Fällen könnten zwei englische Wortsätze lediglich zwei deutsche Wörter als Ausgabe haben, in anderen Fällen sogar vier. Machine Learning Algorithmen erwarten eine feste Ausgabe, die durch eine Klasse von Deep-Learning-Algorithmen überwunden wird, die Sequence-to-Sequence-Learning genannt werden. Die Sequenz-zu-Sequenz-Modelle können einen Eingang und unterschiedliche Ausgangslängen aufweisen. Die Genauigkeit dieser maschinellen Übersetzungssysteme hängt nicht mehr von Regeln oder einfachen Statistiken ab, sondern mehr von der Menge an Trainingsdaten, die zur Verfügung stehen. Jede Sprache kann in jede beliebige Sprache übersetzt werden, wenn man zwei Sprachkorpora als Trainingsdaten mit neuronalen maschinellen Übersetzungen erhält. Nachfolgend finden Sie in Abb. 4.10 eine Benchmark-Studie, die neuronale maschinelle Übersetzungen mit menschlicher Übersetzungsgenauigkeit [18] vergleicht.

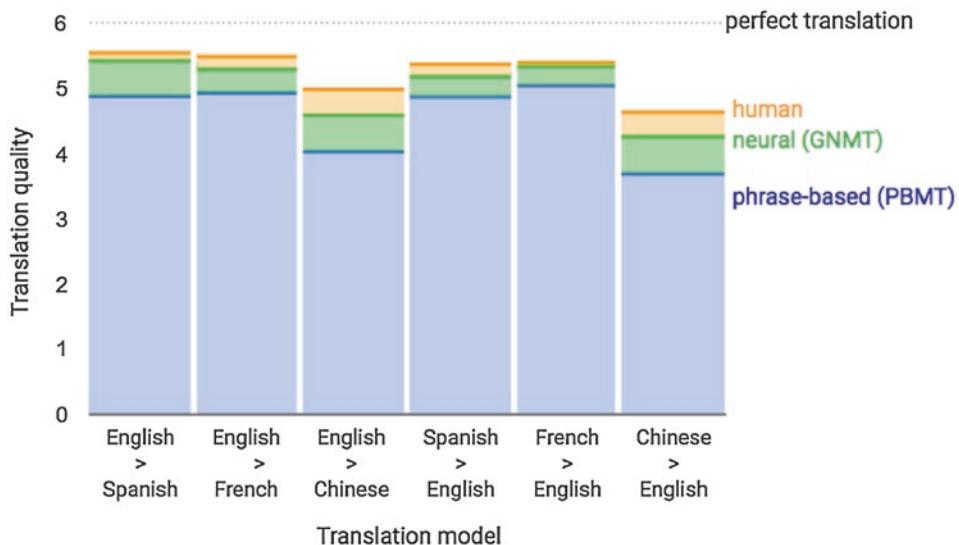


Abb. 4.10 Übersetzungsgenauigkeit bei maschinellen Übersetzungen (Foto: Google)

4. Automatische Generierung von Romanen und fiktionalen Texten: NaNoGenMo (National Novel Generation Month) ist eine offene Veranstaltung zur Generierung natürlicher Sprache, bei der die Teilnehmer verpflichtet sind, automatisiert Romane mit 50.000 Wörtern oder mehr zu erstellen (Zhou und @realsharonzhou). Der Wettbewerb wird bereits seit 2013 ausgerichtet. Die Wettbewerbsteilnehmer haben verschiedene Techniken verwendet, um Romane zu erstellen – wie beispielsweise wortwahrscheinlichkeitsbasierte Markov-Ketten, wiederkehrende neuronale Netzwerke und maschinelle Übersetzungsmethoden. Ein interessantes Werk, das in der Arbeit zitiert wird, ist ein Buch, das mit Jane-Austen-Romanen (Familiendrama) als Input und Arthur Conan Doyle als Output (Detective Genre) erarbeitet wurde.

4.7 Aktuelle Herausforderungen in NLP

Alan Turing sage voraus, dass Gespräche im Jahr 2000 zwischen Mensch und Computer nicht unterscheidbar sein würden. Obwohl die KI individuell in Bezug auf Verständnis und Generierung viel erreicht hat, gibt es noch viel zu tun. Auch 20 Jahre nach dem von Turing avisierten Zeitpunkt sind wir noch weit von dem postulierten Ziel entfernt. Die Gesprächs-KI wird nur dann nahtlos funktionieren, wenn wir folgende Probleme lösen:

- Allgemeiner Kontext oder das „Wissen über die Welt“:** Menschlichen Gesprächen liegt ein inhärentes Verständnis von „Wissen“ und „Logik“ zugrunde. Wenn jemand sagt, dass er einen Sohn hat, dann wissen wir automatisch, dass der Sohn jünger ist als die sprechende Person. Wenn jemand sagt, dass er Vegetarier ist, wissen wir, dass er

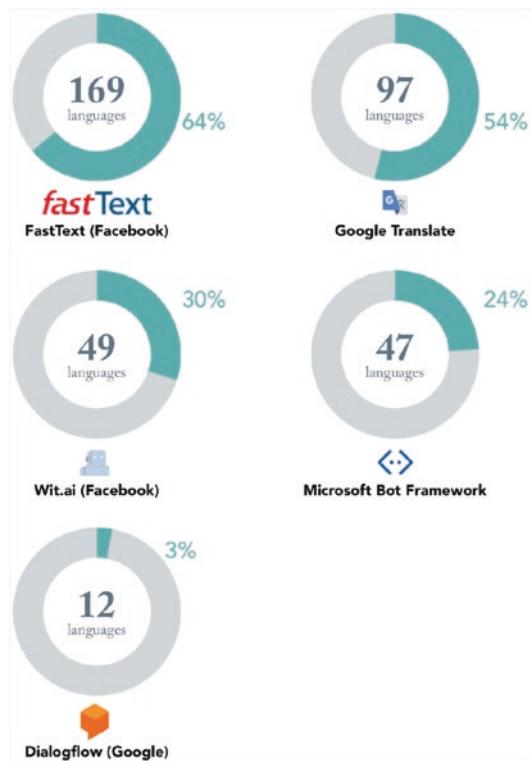
über seine Essgewohnheiten spricht und nicht darüber, was er ist. Diese Schlussfolgerungen sind in menschliche Gespräche integriert, die Maschinen derzeit nicht gesammelt haben.

- „**Vernünftiges Wissen**“ ist eine aktuelle Einschränkung für Computer, um menschenähnliche Gespräche zu führen. Das Alan Institute in Seattle ist eine gemeinnützige Organisation, die erforscht, wie man Computern gesunden Menschenverstand und Logik vermittelt. Ein Problem, das voraussichtlich solange schwer zu lösen sein wird, solange das Wissen nicht strukturiert im Netz verfügbar ist und damit als Referenz für Maschinen nutzbar wird.
- **Mehrdeutigkeit:** Mehrdeutigkeit ist in Sprachen eingebaut und viele Lösungen werden entweder durch den Kontext oder wie im vorherigen Abschnitt durch den gesunden Menschenverstand wahrgenommen. Für Maschinen ist es jedoch schwierig, diese Unklarheiten zu beheben. Nachfolgend finden Sie Beispiele, die die Verwirrung erklären, die durch die natürliche Sprache verursacht werden kann.
 - **Lexikalische Ambiguität:** Wenn das Wort verschiedene Bedeutungen hat. Wikipedia führt dieses Beispiel an: „1916 erkrankte Maurice Ravel an der Ruhr.“ Bedeutet das Wort „Ruhr“ den gleichnamigen Fluss in Nordrhein-Westfalen oder die schwere Durchfallerkrankung?
 - **Mehrdeutigkeit auf Syntax-Ebene:** Der gleiche Satz kann verschiedene Dinge bedeuten. „Marcy hat das Bad für ihre Tochter mit einem rosa Tutu vorbereitet“. Trug Marcy das Tutu? Oder war es ihre Tochter?
 - **Referenzielle Mehrdeutigkeit:** Bezieht sich auf etwas, das Pronomen verwendet. „Eva sprach mit Anna. Sie sagte, du siehst wunderschön aus!“ Wer ist hier wunderschön – Eva oder Anna?
- **Sarkasmus:** Dies ist ein weiterer Bereich, in dem Maschinen typischerweise schlechter abschneiden als ihre „intelligenten“ menschlichen Kollegen. Sarkasmus erfordert nicht nur das Verständnis des Kontextes, sondern in bestimmten Fällen auch Mimik, Betonung und Kultur des Sprechers. NLP-Anwendungsfälle, die sich mit Reviews, Tweets etc. befassen, scheitern in dieser Hinsicht oft. Dies ist ein Bereich, der intensiv erforscht wird, aber aktuell definitiv nicht zu den Stärken von NLP gehört.

4.8 NLP für Nicht-Englisch

NLP wurde intensiv für die englische Sprache erforscht und entwickelt. Dabei sprechen nur fünf Prozent der Weltbevölkerung diese Sprache. Die Anzahl, der für andere Sprachen verfügbaren NLP-Tools, ist relativ begrenzt. Abb. 4.11 zeigt die Anwendung von NLP-Tools für Sprachen von Menschen, die unterhalb der Armutsgrenze leben. Wir müssen davon ausgehen, dass die Qualität von NLP im Vergleich zu Englisch meist deutlich geringer ist. Fehlende verfügbare Datensätze, kaum kommerzielle Gründe und fehlende Anreize für Forscher halten NLP von nicht-englischen Sprachen fern. Ein weiteres großes

Abb. 4.11 Anteil der Menschen, die mit weniger als 1,50 Dollar pro Tag auskommen müssen und deren Sprache von jeweiligen NLP-Tools verstanden werden (Foto: Weltbank, Ethnologue, Webseiten der Software-Anbieter)



Problem für nicht-englische Sprachen ist das „Code-Mixing“. Eine oder mehrere Sprachen sind Teil desselben Satzes. Das Problem wird noch verschärft, wenn Wörter in der Muttersprache transliteriert werden, anstatt ihr eigenes Skript zu verwenden.

4.9 Ausblick

Laut dem Marktforschungsunternehmen Tractica wird NLP bis 2025 auf eine 22 Milliarden-Dollar-Industrie geschätzt.

Mit tonnenweise Forschungsarbeiten, die NLP intelligenter machen, werden Dialoge mit Maschinen in Zukunft weniger von Dialogen unterscheidbar, die wir mit Menschen führen. Durch die qualitativen Fortschritte in allen Bereichen des Natural Language Processing werden zunehmend Interaktionen, die in Unternehmen stattfinden – durch Suche oder menschlich unterstützte Interaktionen – weitgehend von „intelligenten Assistenten“ gesteuert. Das digitalisierte Wissen wächst weiterhin stark an, sodass immer weitere Wissensfelder für das Training von Algorithmen verwendet werden können. Natural-Language-Understanding-Algorithmen werden gleichzeitig immer besser und handlicher, sodass NLU in Kürze in viele Geräte eingebaut werden kann, die wir uns heute nur schwer

vorstellen können. Was wir heute in Form von Anwendungen erleben, ist nur die Spitze eines Eisbergs. In Zukunft werden Bots in unserem Alltag zusammenwachsen und zur Norm werden. Es wird eventuell eine Zeit geben, in der es für Menschen einfacher ist, mit Maschinen zu interagieren als mit anderen Menschen zu sprechen. So wie wir von Bankfilialen auf Geldautomaten und von Fahrkartenschaltern auf Apps umgestellt haben.

Literatur

1. Natural language processing (Definition), hier mit eigener Übersetzung ins Deutsche https://en.wikipedia.org/wiki/Natural_language_processing, zugegriffen 22.11.2019
2. 1 the Road, Wikipedia, https://en.wikipedia.org/wiki/1_the_Road, zugegriffen 22.11.2019
3. The History of Search Engines – A Timeline: <https://www.wordstream.com/articles/internet-search-engines-history>
4. As we may think, Wikipedia, https://de.wikipedia.org/wiki/As_We_May_Think, zugegriffen 21.11.2019
5. Fishkin, Rand: The State of Searcher Behavior Revealed Through 23 Remarkable Statistics, (14.3.2017), <https://moz.com/blog/state-of-searcher-behavior-revealed>, zugegriffen 20.11.2019
6. Fullerton, Laurie: Online Reviews Impact Purchasing Decisions for over 93% of Consumers, Report Suggests, (27.03.2017), <https://www.thedrum.com/news/2017/03/27/online-reviews-impact-purchasing-decisions-over-93-consumers-report-suggests>, zugegriffen 05.11.
7. Allen, Kevin: „Report: 55 Percent of Consumers Complain on Social Media to See Results“, 30.08.2017, <https://www.prdaily.com/report-55-percent-of-consumers-complain-on-social-media-to-see-results/>, zugegriffen 22.11.2019
8. Mindruta, Ruxandra: The Top Social Media Monitoring Tools, (27.09.2019), <https://www.brandwatch.com/blog/top-social-media-monitoring-tools/>, zugegriffen 25.11.2019
9. Zion Market Research. 2019. „Global LegalTech Artificial Intelligence Market Will Reach USD 37,858 Million By 2026: Zion Market Research.“ GlobeNewswire News Room. March 27, 2019. <http://www.globenewswire.com/news-release/2019/03/27/1773829/0/en/Global-LegalTech-Artificial-Intelligence-Market-Will-Reach-USD-37-858-Million-By-2026-Zion-Market-Research.html>
10. Volna, Philip: „Price Wars In The E-Commerce Industry: How Big Data Helps Businesses To Gain Market Share.“ (18.8.2018), <https://www.business2community.com/big-data/price-wars-in-the-e-commerce-industry-how-big-data-helps-businesses-to-gain-market-share-02109204>, zugegriffen 22.11.2019
11. Peiser, Jaclyn: The Rise of the Robot Reporter, (5.2.2019), <https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html>, zugegriffen, 01.11.2019
12. Lever, Rob: Robo-journalism gains traction in shifting media landscape, (10.3.2019), <https://phys.org/news/2019-03-robo-journalism-gains-traction-shifting-media.html>, zugegriffen 23.11.2019
13. World Economic Forum: This is what happens in a minute on the internet, <https://www.weforum.org/agenda/2019/03/what-happens-in-an-internet-minute-in-2019/>, zugegriffen 25.11.2019
14. *The Big Bang Theory* (TV Series 2007-) – IMDb. 2006. <http://www.imdb.com/title/tt0898266/>, zugegriffen 25.11.2019
15. The Beginner’s Guide to SEO: How Search Engines Work: Crawling, Indexing, and Ranking, (01.04.2019), <https://moz.com/beginners-guide-to-seo/how-search-engines-operate>, zugegriffen 25.11.2029

16. Inverted Index – Wikipedia, Wikimedia Foundation, Inc. November 10, 2005. https://en.m.wikipedia.org/wiki/Inverted_index, zugegriffen 25.11.2019
17. Geitgey, Adam: Machine Learning Is Fun Part 5: Language Translation with Deep Learning and the Magic of Sequences, <https://medium.com/@ageitgey/machine-learning-is-fun-part-5-language-translation-with-deep-learning-and-the-magic-of-sequences-2ace0acca0aa> (21.08.2016), zugegriffen 22.11.2019
18. Zhou, Sharon: Has AI surpassed humans at translation? Not even close! (25.07.2019) https://www.skynettoday.com/editorials/state_of_nmt, zugegriffen 22.11.2019



Weshalb sind CUIs und intelligente Assistenten so bedeutend?

5

Zusammenfassung

Der Wettlauf um den führenden, universellen CUI hat längst begonnen. Diese Veränderung wird schneller und massiver ausfallen, als viele heute vermuten. CUIs sind bereits in Lautsprechern, Smartphones, Handys und eigenen Gehäusen (Dots, Pods etc.) vorhanden. Neue Heimelektronik (Licht, Klima, Küchengeräte, Fernseher etc.) kommt ganz selbstverständlich mit Sprachsteuerung auf den Markt. Immer mehr Marketingabteilungen großer Unternehmen entdecken die Macht der Sprache und testen Bots sowie Assistenten. Nutzerwünsche werden kontextbasiert erkannt und Algorithmen analysieren die Absichten der Kunden. CUIs erkennen in Kombination mit Sensorik Bedürfnisse ihrer Nutzer, noch bevor sie ausgesprochen werden. Conversational User Interfaces werden nicht nur im Handel, sondern in allen Bereichen des Lebens eine bedeutende Rolle einnehmen.

Wenn wir uns an guten Service gewöhnt haben, wollen wir nicht mehr zurück. Wenn wir gelernt haben, dass Kleidung, Lebensmittel und technische Geräte bis an die Haustür geliefert werden, warum sollten wir dann Zeit mit der Fahrt in Geschäfte sowie Anstehen an der Kasse verbringen? Wenn wir erfahren haben, welchen Unterhaltungswert Binge-Watching der neuen Lieblingsserie beim Streaming-Anbieter macht, warum sollte man dann noch auf feste Sendetermine warten und Werbeunterbrechungen ertragen. Wenn man weiß, wie leicht ein Zugticket online gebucht ist, warum sollte man sich an einem Schalter anstellen? Es gibt dutzende Dinge in unserem Alltag, die aufgrund der Digitalisierung einfacher, schneller, bequemer geworden sind. Es gibt wohl kaum einen Nutzer, der wieder zum „Alten“ zurück möchte. Anbieter, die diese neu entstandene Serviceerwartung nicht befriedigen, geraten schnell ins Abseits.

Genauso wird es mit CUIs verlaufen. Wenn Nutzer sich erst einmal daran gewöhnt haben, jederzeit und überall ihre Wünsche zu äußern, die dann auch direkt umgesetzt werden, gibt es kein Zurück mehr. Kein Durchklicken von Links auf unübersichtlichen Webseiten, kein Ausfüllen von Bestellformularen bei Dienstleistern. Das Versprechen lautet: Du musst nie wieder suchen und dann mithilfe der möglicherweise unter Mühen gefundenen Informationen die beabsichtigte Aufgabe selbst erledigen. Der Nutzer äußert, was er möchte, und die Maschine versteht und setzt es um.

So wie das Smartphone kein kleiner Computer ist und Apps keine vertikal angeordneten Webseiten, so sind auch CUIs ein vollkommen neues Feld. Es ist eine massive Disruption, deren Tragweite von vielen noch nicht richtig eingeschätzt wird. Die Entwicklung des Smartphones von der ersten Markteinführung der Geräte bis zur Reife des Marktes hat ungefähr fünf Jahre benötigt. Selbst wenn man davon ausgeht, dass die Komplexität beim Ausbau von CUIs und Assistenten um einiges höher ist, als beim Ausrollen des Smartphone-Ökosystems, ist es andererseits heute so, dass die Entwicklungsgeschwindigkeit immer schneller wird. Darum können wir davon ausgehen, dass in fünf Jahren die Nutzung von CUIs und intelligenten Assistenten einen Reifegrad erreicht haben wird, der zu einem neuen Service-Angebots-Standard geworden ist, hinter dem kein Unternehmen und keine Institution mehr zurückbleiben will. Mit anderen Worten: in fünf Jahren, also voraussichtlich bereits 2025, ist die CUI-Fähigkeit von Unternehmen und Institutionen ein Muss und diejenigen Marktteilnehmer, die keine Voice- oder Semantic-Strategie entwickelt haben, werden geschwächt im Wettbewerb auftreten. Diese Unternehmen werden existenzielle Probleme bekommen, genau wie jene, die die Veränderungsintensität des Internets unterschätzt oder den Umstieg vom Desktop auf Mobile verschlafen haben.

5.1 Warum jetzt? Und weshalb so schnell?

Wenn eine Transformation dieser Größenordnung stattfindet, ist das Timing entscheidend. Wer erwägt, in diesen neuen Interaktionskanal zu investieren, muss wissen, wann Fantasie in Realität übergeht und wann man die aufkommende Technologie in einen Geschäftsvorteil verwandeln kann. Ist man zu früh dran, verschwendet man Mühe mit einer unreifen Technologie. Ist man zu spät, bleibt man hinter den Konurrenten zurück und kann unter Umständen nicht mehr aufholen. Der Moment für intelligente Assistenten ist mittlerweile definitiv gekommen.

Die Gründe dafür sind, dass jetzt die folgenden drei Voraussetzungen erfüllt sind:

- **Erstens:** Die Nutzerwünsche sind bekannt. Computersysteme in Unternehmen haben viele Informationen über die Nutzer und ihre Welt. Basierend auf dem Wissen, wer anruft, wo die Verbraucher sind, welche Tageszeit es ist, wie das Wetter ist und vielen anderen kontextbasierten Informationen, kann eine intelligente Software erkennen, wonach die Nutzer suchen oder voraussichtlich in Kürze suchen werden. Dabei sollte man sich nicht von den häufig verwirrend dümmlichen Targeting-Ads in Sozialen Medien

irritieren lassen. Oder das unverständliche Re-Targeting von Produktanzeigen als Maßstab anlegen, deren Inhalt man gerade gekauft hat. Hinweise auf die Qualität und Menge der Daten, die über uns als einzelne oder Gruppe verfügbar sind, erhält man bei jeder Google-Suche auf einem Smartphone, bei denen Suchergebnisse in der Umgebung dargestellt werden, die nach aktuellen Öffnungszeiten gerankt sind. Das ist schon ganz schön „magic“ im Vergleich zu den Suchergebnissen, die man noch vor kurzem erhalten hat.

- **Zweitens:** Die automatisierte Analyse der Kundenabsicht ist möglich, wenngleich aber noch schwierig. Will man verstehen, was Kunden wirklich wollen, reicht es nicht aus, einfache Entscheidungsbäume zu erstellen, die unterschiedliche Entscheidungen treffen, basierend auf „wenn das, dann das“ -Regeln. Wie jeder weiß, der jemals versucht hat, die Absicht eines anderen Menschen herauszufinden: Man kann sich leicht irren. Zunächst muss das System, das den intelligenten Assistenten antreibt, über Merkmale des Unternehmens antrainieren und plausible Modelle erarbeiten, die abbilden, wie Kunden mit dem Unternehmen üblicherweise in der Vergangenheit interagiert haben. Dann benötigt dieses System riesige Datenmengen, mit denen es üben kann. Schließlich kann das System selbstständig einen „intelligenten“ Algorithmus entwickeln und verwenden, damit der intelligente Agent eine gute Einschätzung darüber abgeben kann, was der Kunde verlangen wird. Selbst wenn dieser Algorithmus effektiv ist, ist er nicht statisch; da sich Geschäftsbedingungen, Produkte und Preise ändern, muss sich auch der Algorithmus, der den intelligenten Assistenten unterstützt, ändern, um Schritt zu halten. Die Abbildung von Kundenverhalten in Modellen und die Erprobung und Optimierung dieser Modelle ist heute technisch möglich, und die Erfahrung im Bearbeiten der Machine-Learning-Algorithmen ist ausreichend vorhanden.
- **Drittens:** Die Hardwarevoraussetzungen sind sowohl bei Nutzern, als auch bei Unternehmen vorhanden. Nutzer sind heute mit einer Vielzahl von Geräten, die ständig mit dem Internet verbunden sind, unterwegs. Sie speisen die Cloud mit immer neuen Daten und sind über diese Endgeräte für Unternehmen erreichbar und ansprechbar. Unternehmen wiederum können die großen Datenmengen wirtschaftlich plausibel speichern und verarbeiten.

Wenn also ausreichend Daten vorhanden sind und das Wissen darüber vorliegt, wie mit diesen Daten robuste Anwendungen unterhalten werden können, sowie die Hardware-Voraussetzungen gegeben sind, um diese Anwendungen 24/7 zu betreiben, dann beginnt eine Dynamik, an deren Anfang wir uns gerade befinden.

Die Möglichkeit, die Kundenabsicht zu erfassen, verändert alles. Genau wie einst das Internet, wird das Aufkommen von KI-basierten intelligenten Assistenten, die herausfinden können, was Sie wollen, verändern, wie Unternehmen und ihre Kunden interagieren. Ein intelligenter Agent, der Ihre Absicht voraussehen kann, kann Ihre Probleme schneller lösen. Er kann Erkenntnisse liefern, die nicht nur darauf basieren, was Kunden tun, sondern auch, was sie zukünftig wollen. Er erstellt eine Roadmap zu besseren Kundenerlebnissen, die langfristige Loyalität schafft. Und der Assistent kann genau den Moment

herausfinden, in dem Sie bereit sein könnten, eine tiefere Beziehung zu Unternehmen und Institution einzugehen, die für Unternehmen und Institutionen wirtschaftlich attraktiv sind. Dies sind keine kleinen, inkrementellen Änderungen. Die Autoren Harley Manning und Kerry Bodine haben bereits 2012 in einem Buch [1] herausgearbeitet, wie Unternehmen, die Verbesserungen an der Kundenzufriedenheit vornehmen, breite und dauerhafte Vorteile erzielen. Unternehmen mit herausragend positiver Kundenservicewahrnehmung wachsen schneller und haben einen höheren Börsenwert als diejenigen, die diese Fähigkeiten nicht im gleichen Maße entwickelt haben. Dieser Effekt wird in einer Zeit deutlich zunehmen, in der das Serviceversprechen an Nutzer viel weitreichender ist, als dies in den Zeiten der ersten Websites und einem rudimentären Internet Anfang der 2010er-Jahre der Fall gewesen ist.

Verbesserungen in diesem Bereich sind aber auch herausfordernd und schwierig. Richtig gestaltete intelligente Assistenten können alle kundenorientierten Disziplinen verbessern, einschließlich Vertrieb, Kundenservice und Marketing. Sie können auf die vielfältigen, oft schlecht integrierten Kundensysteme zugreifen und als Frontend für diese fungieren, die in jedem modernen Unternehmen existieren. Sie sind für Kunden einfach zu handhaben und reduzieren die Zeit, die benötigt wird, um einen Verkauf abzuschließen oder ein Problem zu lösen. Selbst wenn die Interaktion mit einer realen Person notwendig ist, um ein Problem zu lösen, können intelligente Assistenzsysteme die Menschen unterstützen, indem sie ihnen die bestmöglichen Informationen zur Verfügung stellen, damit sie den Kunden helfen können. Intelligente Assistenten stellen einen Wendepunkt dar – eine seltene Gelegenheit, Kosten zu senken und gleichzeitig das Kundenerlebnis zu verbessern. Die Unternehmen, die diese neue Art der Geschäftstätigkeit annehmen, werden einen deutlichen Vorsprung gegenüber Wettbewerbern aufbauen, die noch nicht verstanden haben, welchen Unterschied diese neue verbraucherorientierte Technologie machen wird.

5.2 Der Wettkampf hat längst begonnen

Die beschriebenen Entwicklungen werden zu einer starken Veränderung des gesamten digitalen Ökosystems führen – mit Chancen für bestehende aber auch vollkommen neue Marktteilnehmer. Allerdings gibt es auch Risiken für diejenigen, die sich an diese Veränderung zu spät oder zu langsam anpassen. Das erkennen auch die großen Tech-Unternehmen wie Facebook, Amazon, Microsoft, Google und Apple (siehe Abb. 5.1). Sie investieren in nie gekannter Intensität in dieses Themenfeld, weil sie wissen, dass es letztlich um ihre Existenz geht.

Das Rennen um die Eroberung von Sprache und Conversational User Interfaces ist eröffnet. Facebook, Amazon, Microsoft, Google und Apple (FAMGA) haben sich alle als Big Player im Themenfeld der natürlichen Sprache etabliert. Alle investieren in Start-ups, Forschung und Entwicklung, testen die Entwicklungen bei ihren Nutzern – und wetten, dass Sprache die nächste große Plattform für Verbraucher und Unternehmen werden wird. Bevor wir uns den eher allgemeinen Entwicklungsdynamiken zuwenden, wollen wir betrachten, wie die großen westlichen Unternehmen derzeit aufgestellt sind.

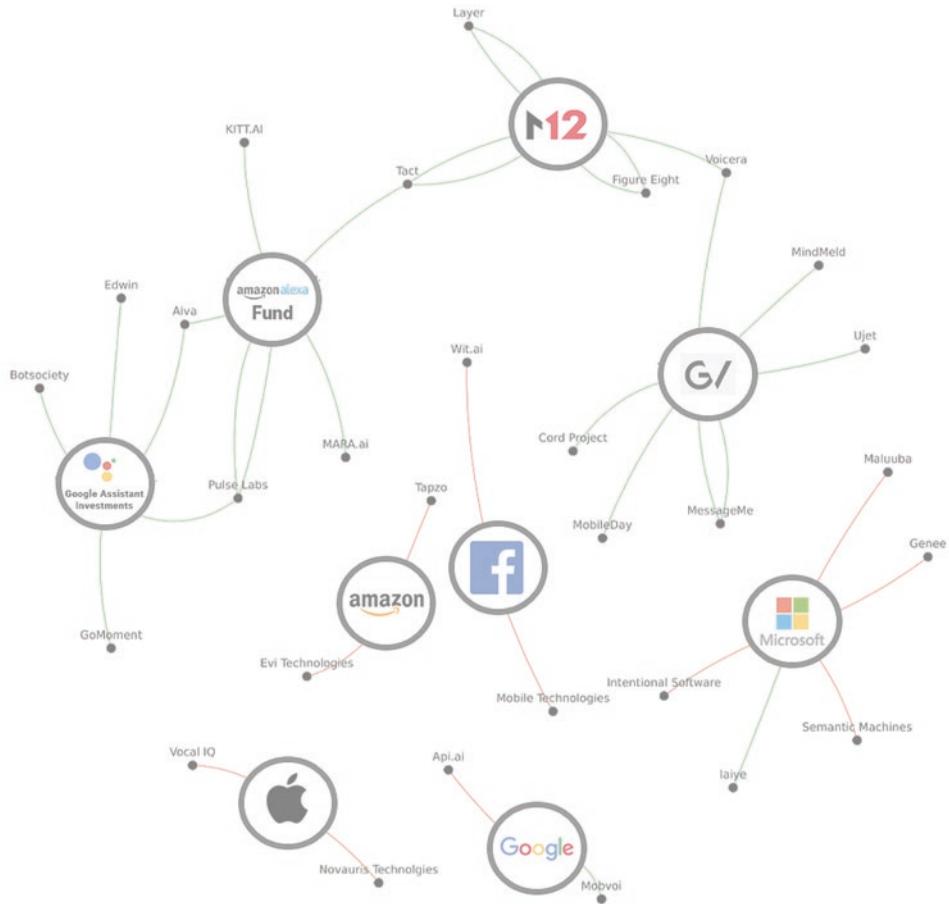


Abb. 5.1 Investment-Aktivitäten von Facebook, Apple, Microsoft, Google, Amazon 2013 bis 2019.
(Foto: CB Insights)

5.3 Google knüpft ein weitgespanntes Netz

Ende 2019 verfügt Google über ein Produktpotfolio mit natürlchsprachigen Funktionen in mehr als 30 Sprachen – einschließlich der Fähigkeit, Nutzer zu verstehen, die mehr als eine Sprache gleichzeitig sprechen und in ihren Äußerungen und Dialogen zwischen verschiedenen Sprachen hin- und herspringen.

Sowohl Google als auch Apple genießen den Vorteil, mit Android und iOS zwei der am weitesten verbreiteten Smartphone-Betriebssysteme zu orchestrieren. Diese liefern täglich nicht nur eine Fülle von Daten darüber, wie Kunden Produkte verwenden, sondern diese Benutzeroberfläche ermöglicht es den Unternehmen, einfache Integrationen zu erstellen, die eine nahtlose Verwendung zwischen Zuhause, Büro und Auto ermöglichen. Mit all diesen Daten hat Google bereits frühzeitig erkannt, dass Kunden ihre Stimme für

Aufgaben nutzen, die über das Tätigen von Anrufen und das Fragen nach dem Weg deutlich hinausgehen. Die Anzahl der Sprachschanfragen steigt stetig an. Google verwendet AI, um die Unterschiede zwischen der Herangehensweise der Nutzer an die Textsuche und der Sprachsue zu verstehen und um die Absichten zu bestimmen, die Nutzer bei ihrer Suche wirklich haben.

Im Wettkampf der besten Natural-Language-Processing-Fähigkeiten hat Google insbesondere gegen Amazon und Apple bereits heute eine führende Position. Das betrifft sowohl Suchanfragen, als auch Aufgabenanfragen. Durch geschickte Akquisitionen baut Google diesen Vorsprung derzeit weiter aus.

Vor kurzem hat Google einen neuen Risikofonds für Investitionen in Voice- und Sprachtechnologien aufgelegt. Google Assistant Investments [2] hat bisher knapp 15 Investitionen getätigt. Erkennbar ist dabei ein Investmentfokus im Bereich Gesundheits- und Gastgewerbe. Insbesondere an zwei Investitionen kann man die Stoßrichtung von Google bei Sprache in den Industriesegmenten erkennen: AVIA Health (siehe Kap. 11), in das sowohl Google Assistant Investments als auch das Amazon Pendant Amazon Alexa Fund investiert haben. AVIA ist ein patientenorientiertes Sprachbetriebssystem, das in Krankenhäusern und Pflegeheimen zum Einsatz kommt. Über Smartspeaker können Patienten mithilfe natürlicher Sprache Unterstützung anfordern, und medizinische Fachkräfte können über mobile Apps auf diese Anfragen reagieren und diese nachverfolgen. Google Assistant hilft auf diese Weise Patienten, aber auch den Krankenhäusern, bei der Betriebsorganisation. Ein weiteres Investment aus dem Google Assistant Investment Portfolio ist NOVANT Health-System. Über Sprachsteuerung können Patienten auf Informationen zugreifen, einschließlich Wartezeiten im Büro, Terminkalender und Anweisungen. Antworten und Informationen werden an das Telefon des Benutzers gesendet. Diese Investitionen stehen im Zusammenhang mit dem zunehmenden Engagement von Google im Gesundheitswesen und bieten Patienten über Google Home und Google Assistant einen weiteren Einstiegspunkt.

Ein weiterer Investitionschwerpunkt ist die Hospitality-Branche. GoMoment bietet eine neue Art, mit dem Concierge eines Hotels ins Gespräch zu kommen. Es ist ein intelligenter Chatbot, der Hotelgäste begrüßt und Fragen beantwortet. Es kann auch die Zufriedenheit der Gäste abfragen und die Wartezeit von Gästen beim Anruf des Zimmerservice für Hoteldienstleistungen verkürzen. GoMoment stellt gleichzeitig den „Dolmetschermodus“ von Google bereit. Im Rahmen von Pilotprogrammen in einigen Hotelketten können zwei Personen, zum Beispiel ein ausländischer Gast und ein Hotelmitarbeiter an der Rezeption, ihre Konversation in Echtzeit führen. Die Konversation wird sowohl über einen Smartspeaker von Google Home als auch über die Google-Assistant-App übersetzt und auf einem Tablet angezeigt, sodass beide Parteien miteinander kommunizieren können.

Wie geschickt Google dabei seine verschiedenen Produkte mittlerweile zusammenfügt, wird beispielsweise hier deutlich: Mit einer personalisierten Google-Startseite im Hotelzimmer, das in das Gastgewerbesystem GoMoment integriert ist, können Gäste Veranstaltungstickets erwerben, den Zimmerservice bestellen, Wetterinformationen abrufen und ihre Reiseinformationen über Sprache abrufen. Dabei verschwimmt für den Gast weitestgehend der Unterschied zwischen der zuhause üblichen Google-Startseite und der Hotel-Startseite.

5.4 Amazon: Marktführer bei Smartspeakern

Obwohl Google beeindruckende Fortschritte gemacht hat, ist Amazon immer noch der unbestrittene Marktführer des Segments der Smartspeaker mit Alexa. Natürlich hat sich Amazon darauf konzentriert, sein Sprachangebot durch E-Commerce zu verbessern. Hierbei hat Amazon einen echten Vorteil gegenüber seinen FAMGA-Konkurrenten, da jedes Alexa-Gerät direkt mit einem der weltweit größten Online-Shops verbunden ist. Amazon erschließt die Schnittstelle zwischen Sprache und E-Commerce über Amazon Choice – ein Label, das verschiedenen Produkten auf der Website verliehen wird und das Amazon als „hoch bewertete, preisgünstige Produkte, die sofort lieferbar sind“ bezeichnet.

Die OC&C-Strategy-Consultants weisen darauf hin, dass Artikel mit der Amazon-Choice-Bezeichnung im sprachgesteuerten Amazon-Shop in der Regel einen dreifachen Umsatzzuwachs gegenüber Artikeln ohne Auswahlstatus erzielen. Vielleicht sogar noch wichtiger: 85 Prozent der Verbraucher entscheiden sich letztendlich für das von Amazon vorgeschlagene Produkt. Für Amazon bedeutet das Potenzial für Marken, in das von Amazon Choice vorgeschlagene Produkt bei Alexa einzusteigen, gleichzeitig wird erneut deutlich, wie noch viel mächtiger die Gatekeeper-Funktion eines Unternehmens wie Amazon im Umfeld von CUIs und Assistenten werden wird.

Untersuchungen haben jedoch ebenfalls gezeigt, dass derzeit nur 30 Prozent der Smartspeaker-Besitzer sie zum Einkaufen verwenden. Damit kann Amazon nicht zufrieden sein, denn nur zwei Prozent der Echo-Besitzer haben 2018 versucht, Artikel über Alexa zu kaufen. Das heißt aber nicht, dass es kein Potenzial gibt. Amazon arbeitet daran, sich mit seinem eingebauten Vorteil auf der Einzelhandels- und E-Commerce-Seite zu positionieren, um den Voice-Shopping-Bereich zu dominieren. Amazon plant dies, indem es mehr intelligente Produkte in Haushalte bringt, die mit Alexa kommunizieren. Zum anderen arbeitet Amazon mit Microsoft zusammen, einem Unternehmen, das mit Apple und Google technologisch konkurrieren kann und Amazon mit Windows ein Betriebssystem zur Verfügung stellt, mit dem es zusammenarbeiten kann. Darüber hinaus strebt das Unternehmen Partnerschaften mit Packaged-Goods-Unternehmen an, damit diese für Alexa-Funktionen werben. Natürlich strebt Amazon auch eine Expansion in vollkommen neue Bereiche an. Das Unternehmen war außerhalb des eigenen Kerngeschäfts immer im Nachteil gegenüber Google und Apple, aber es versucht, dies zu ändern, und plant neue sprachaktivierte Geräte, die dem Unternehmen einen größeren Vorteil bei der Datenerfassung verschaffen würden. Bei dem im September 2019 durchgeführten Alexa Devices Event wurden nicht nur eine ganze Reihe neuer und auf Klangqualität optimierte Smartspeaker vorgestellt, sondern mit Echo-Buds auch ein ernst zu nehmender Wettbewerber zu den kabellosen Kopfhörern von Apple. Selbstverständlich ist Amazons sprachgesteuerter Assistent Alexa über die Kopfhörer ansprechbar. Neben Kopfhörern stellt Amazon auch mit Echo-Frames ein Brillengestell vor, das Alexa ansprechen lässt und mit Echo-Loop einen Fingerring, der Mikrofone enthält, die Nutzer direkt Alexa ansprechen lassen.

„Alexa, lass uns eine Ausfahrt zusammen machen!“

Wie man an diesen Produktvorstellungen erkennt, arbeitet Amazon intensiv daran den Nachteil aufzuholen, der dadurch entstanden ist, dass Alexa nicht Bestandteil eines Betriebssystems ist. Während Google und Apple über Telefonschnittstellen mit Assistenten verfügen, die für den mobilen Einsatz prädestiniert sind, ist Alexa ein reines Heimprodukt – bis jetzt.

Mit der kürzlich erfolgten Einführung von Echo Auto (mit einer Million Vorbestellungen) versucht das Unternehmen, diejenigen Kunden anzusprechen, die mehr Optionen für die Freisprechfunktion von Sprachassistenten wünschen. Mit Echo Auto können Benutzer auf alle Alexa-Fähigkeiten zugreifen. Sie können Anrufe, Wegbeschreibungen, Podcasts, Musikwiedergabelisten und die Steuerung von Smart-Home-Geräten freihändig tätigen. Umfragedaten zeigen, dass die Verwendung von Sprachassistenten in Autos bereits sehr beliebt ist. Daher ist dies eine natürliche Ergänzung für Big Tech. Amazon befindet sich derzeit in Gesprächen mit großen Autoherstellern, darunter BMW, Toyota und Ford, um Alexa in ihre Infotainmentsysteme zu integrieren. Eine Reihe anderer Hersteller wie Infinity, Jaguar und Mercedes-Benz erweitern neuere Modelle um die Kompatibilität mit Alexa-Fähigkeiten. Ebenfalls auf dem Echo Devices Event im September 2019 verkündete Amazon eine offensichtlich groß angelegte Partnerschaft mit GM, das in seine Fahrzeuge Amazons Alexa nicht nur als Add-On zum bestehenden Bediensystem einsetzt, sondern über Echo Auto auch die Kernfunktionen des Fahrzeugs steuerbar machen wird.

Das smarte Hausgerät ist da

Möchten Verbraucher mit ihren Haushaltsgeräten sprechen? Amazon glaubt, die Antwort lautet Ja. Amazon hat das Amazon Connect Kit (ACK) entwickelt, mit dem Entwickler auf die APIs für Alexa-Heimkompetenzen zugreifen können, um Tausenden von Heimgeräten und kleinen Geräten neue und benutzerdefinierte Funktionen hinzuzufügen (siehe Abb. 5.2).

Die Umstellung auf sprachaktivierte Smart Appliances zeigt das Potenzial für eine Alexa-Erweiterung. Jeff Bezos hat in seinem Q2 2018 Earnings Call deutlich gemacht, wie wichtig Stimme für die Zukunft des Unternehmens ist.

„Mittlerweile bauen Zehntausende Entwickler in mehr als 150 Ländern neue Geräte mit dem Alexa Voice Service, und die Anzahl der Alexa-fähigen Geräte hat sich im vergangenen Jahr mehr als verdreifacht. Unsere Partner entwickeln eine Vielzahl neuer Alexa-fähiger Geräte und Erlebnisse“, so Bezos. Amazon gibt an, 28.000 verschiedene Smart-Home-Geräte von 4500 Marken mit Alexa-Funktionen ausgestattet zu haben.

Der Umsatz mit intelligenten Geräten wird voraussichtlich zunächst langsam steigen. Nicht zuletzt, weil Verbraucher Kühlschränke, Waschmaschinen, aber auch TV-Geräte meistens mehrere Jahre nutzen, ehe sie die alten durch neue Geräte ersetzen. Dies gilt noch mehr für Heizungsanlagen und andere Geräte der Heimsteuerung. Wichtige Appliance-Marken wie LG, Samsung, Whirlpool, Kenmore, Bosch und GE bieten heute große Portfolios intelligenter Haushaltsgeräte an, die sowohl mit Alexa als auch mit Google Home verbunden sind. Eine intelligente Mikrowelle ist bereits über AmazonBasics

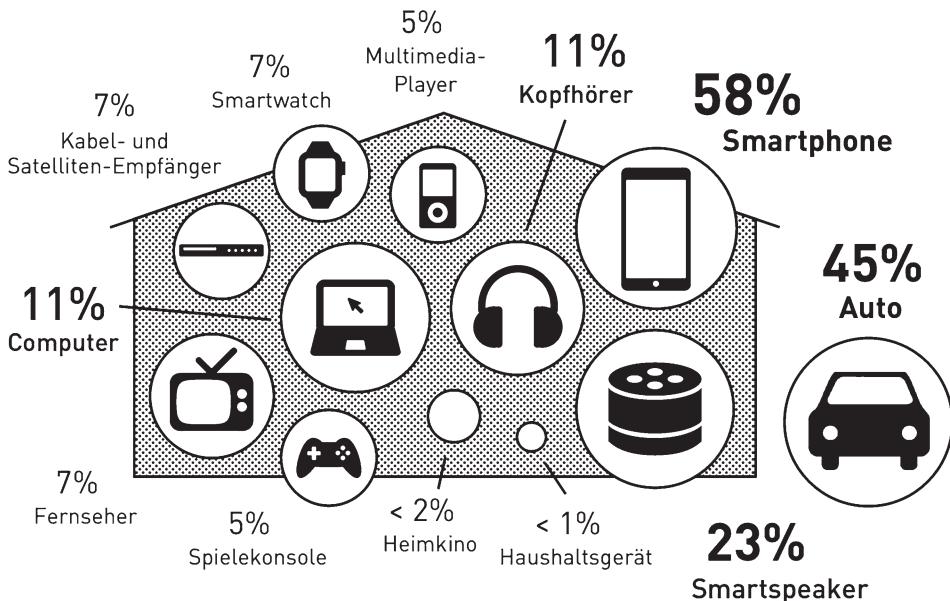


Abb. 5.2 Nutzung von Voice Assistants nach Endgeräten in den USA (09/2018)

erhältlich. Neue Kompetenzpartnerschaften mit MOEN und iRobot ermöglichen es Benutzern, Duschen einzuschalten und einen Roomba-Staubsaugerroboter über Sprachanweisungen zu aktivieren. Amazon nutzt Amazon Connect Kit (ACK) als Einstieg, um auf ein vollständig integriertes Smart Home zuzugreifen, das mithilfe von Alexa sprachgesteuert wird.

5.5 Apple: Siris Frühstart

Im Jahr 2011, als Apple Siri zum ersten Mal auf den Markt brachte, läutete es – lange vor Google und Amazon eine neue Welt sprachaktivierter Assistenten ein. Eigentlich der richtige Schritt, um diesen Markt nachhaltig zu beherrschen. Allerdings hat Apple erst vergleichsweise spät erkannt, dass es die Smartspeaker sind, die Nutzer an das Phänomen Sprachsteuerung heranführen. Der späte Einstieg in das Smartspeaker-Rennen, mit dem im Vergleich zu den offensichtlich subventionierten Konkurrenzprodukten, recht hochpreisigen HomePod, hat Apple, zumindest momentan, auf einen hinteren Platz zurückgeworfen.

Branchenanalysten nennen einige schwerwiegende Mängel von Siri auf dem HomePod im Vergleich zu seinen größten Konkurrenten – er ist nicht so vielseitig wie Alexa von Amazon oder so „intelligent“ wie das KI-System von Google. Für engagierte Apple-Benutzer spielen diese Probleme keine Rolle. Aber für diejenigen, die nicht voll mit dem Apple-Universum verbunden sind, ist es schwierig, alle Siri-Angebote voll

auszuschöpfen. Ein Bereich, in dem sich der HomePod auszeichnet, ist Privatsphäre und Sicherheit. Die Übernahme von Silk Labs im Jahr 2018, die eine auf dem Gerät installierte AI-Betriebssoftware entwickelt, die die Cloud-Verarbeitung überflüssig macht, spricht dafür, dass Apple versucht, durch Betonung von Sicherheit und Privatsphäre die Märkte zu überzeugen.

5.6 Offline ist das neue Online

Sicherheit und Datenschutz werden für Sprachanwender immer wichtiger. Apple hat mehrere Patente angemeldet, die sich auf Sprache und Sicherheit konzentrieren. Das erste Ziel ist offensichtlich, Siri offline per Spracheingabe zu aktivieren, auch wenn keine Internetverbindung besteht. Bereits heute werden beim Apple HomePod keine Gespräche aufgezeichnet, da die meisten Aktionen nicht über die Cloud, sondern im Gerät selbst ausgeführt werden. Hier verfolgt Apple einen anderen Ansatz als seine Hauptkonkurrenten Google und Amazon, die Interaktionen mithilfe der Cloud auf Unternehmensservern verarbeiten und Sprachverarbeitung on device eher aus Gründen der Performance, als aus Gründen des Datenschutzes betreiben. Nicht erst seit den (vermeintlichen) Skandalen um abgehörte Gespräche, wollen Verbraucher wissen, welche Daten erfasst und gespeichert werden. Apple ist bemüht, Möglichkeiten zur Verwendung der Sprachaktivierung im Gerät selbst zu entwickeln. Offensichtlich will Apple ein System schaffen, bei dem nicht nur die Spracheingabe direkt auf dem Gerät stattfindet, sondern auch die Sprachausgabe über das Gerät des Benutzers und nicht über Apples Server erfolgt. Das Gerät wäre in der Lage, die Anforderung auf der Grundlage vorgegebener Aufgaben auszuführen, die im Betriebssystem gespeichert sind. Benutzer hätten die Möglichkeit, Siri sowohl online als auch offline zu betreiben.

Auch im Bereich der biometrischen Authentifizierung positioniert sich Apple. „Voice prints“ sollen es Siri ermöglichen, mehrere Benutzer zuverlässig durch Stimmerkennung zu identifizieren und zu unterscheiden (siehe Abb. 5.3). Eine praktische Anwendung wäre hier der HomePod, mit dem mehrere Benutzer individuelle Einstellungen auf einem Gerät vornehmen können. Diese Funktionen sind in Ansätzen auch bereits in Amazons Alexa und Google Assistant integriert, allerdings funktioniert die Identifikation einzelner Nutzer bei Amazon und Google derzeit noch so schlecht, dass eine echte Zuordnung von Nutzern auf verschiedene Kundenkonten nicht möglich ist.

5.7 Conversational Productivity Tools – Apple und Microsoft konkurrieren wie in alten Desktop-Zeiten

Wenn große Unternehmen, wie Google, Amazon und Apple eine neue Technologie zum zentralen Bestandteil ihrer jeweiligen Betriebssysteme machen und mithilfe der Technologie in bislang nicht bearbeitete Produktbereiche wie Haushaltsgeräte, Autos und Haussteuerung vordringen, erkennt man die Wucht die von dieser neuen Technologie

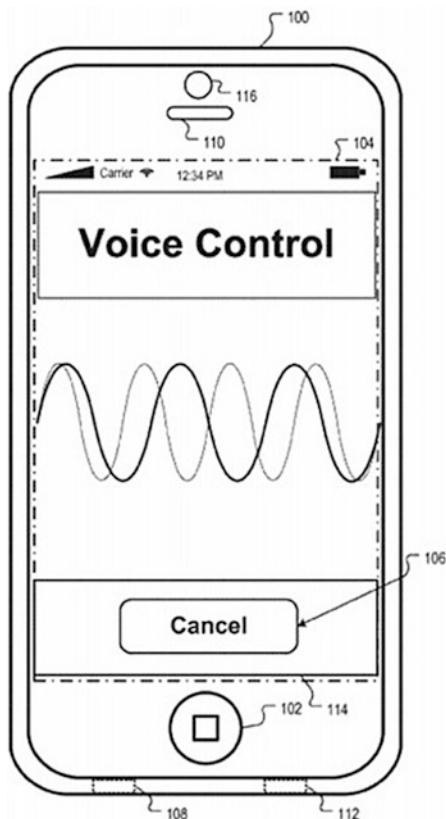


FIG. 1

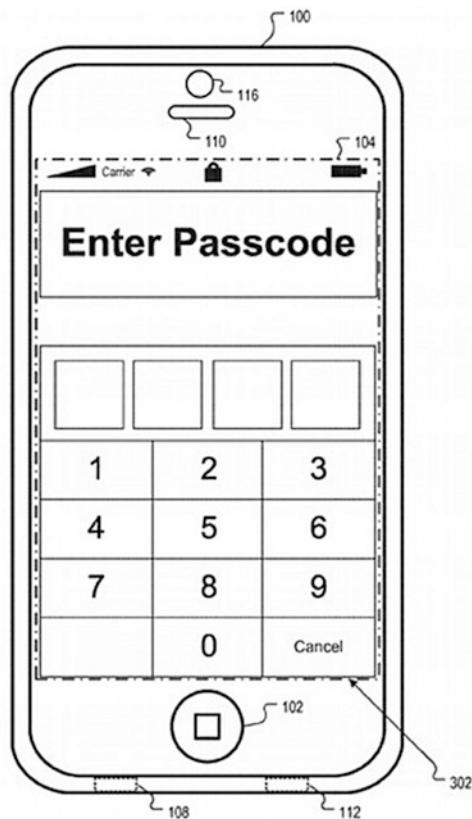


FIG. 3

Abb. 5.3 Zugangssperre per Stimme auf Apples iPhone freischalten (Foto: Apple)

ausgeht. Doch nicht nur zu Hause und bei Privatanwendern wirkt diese Veränderungsdynamik. Auch im professionellen Umfeld gibt es große Veränderungen. Apple beispielsweise möchte, dass Sie Ihre Stimme verwenden, um Dinge zu erledigen. Der Technologieriese arbeitet mit Salesforce zusammen, um die Apple-Sprachtechnologie in Apps zu integrieren.

Benutzer können Siri für Updates und Anfragen verwenden, anstatt zwischen Systemen zu navigieren. Viele der auf Unternehmensservern abgelegten Informationen bleiben ungenutzt, da es für Mitarbeiter von Unternehmen zu unkomfortabel ist, diese Informationen zu finden und in die Entscheidungsprozesse einzubauen. Wenn es unzählige Klicks im ERP-System von SAP oder Oracle bedarf, um die Anzahl der Nutzer zu erfragen, die gleichzeitig in Europa wohnen und häufig in die USA reisen, dann unterbleibt eine solche Analyse im Zweifel. Sobald es den Mitarbeitern aber möglich ist, eine einfache, natürlichsprachige Anfrage an das System zu schicken und ohne mühsames Suchen in den Datenhierarchien die gewünschten Informationen zu erhalten, dann ändert dies auch im

professionellen Zusammenhang einiges. Salesforce ist eines der beliebtesten Tools auf Unternehmensebene, das in Vertrieb und Marketing verfügbar ist. Außerdem verfügt es über eine Armee von Softwareentwicklern von Drittanbietern, die in der Lage sind, ihre Back-End-Technologie zu nutzen und bessere Möglichkeiten zur Integration zu finden. Apple-CEO Tim Cook sagte gegenüber Reuters über die Partnerschaft: „Wenn man sich das Unternehmen im Allgemeinen ansieht, wurde Sprache nicht so häufig wie beim Endverbraucher eingesetzt. Wir verändern die Art und Weise, wie Menschen arbeiten darüber, Dinge zum Besseren zu verändern“ [3].

Obwohl Microsoft ein relativ früher Einstieg in das Smart-Speaker-Rennen mit Cortana gelang, hat es Probleme, im Konsumenten-Umfeld mit Google und Amazon mitzuhalten. Offenbar hat sich Microsoft von der direkten Konkurrenz abgewendet und sucht stattdessen nach Integrationspartnerschaften. Der Softwareentwickler hat sich mit Amazon zusammengetan. Windows-10- und Alexa-Benutzer verfügen nun über systemübergreifende Funktionen. Cortana-Benutzer können über Windows 10 mit Alexa kommunizieren und Alexa-Benutzer können auf Cortana zugreifen, um E-Mails in Outlook vorlesen zu lassen. Dies ist eine strategische Partnerschaft der beiden Technologie-Giganten, die sich auch durch ihre Firmensitze im Großraum Seattle nahe sind. Microsoft wird eine neue Zielgruppe potenzieller Kunden gewinnen, die eine Cortana- und Alexa-Integration als hervorragende Produktivitätsoption betrachten. Insbesondere mit Potenzial auf Unternehmensseite, bei dem Alexa-Kenntnisse für einfache arbeitsbezogene Aufgaben eingesetzt werden könnten. Unternehmensanwender setzen auf Microsoft Cortana, um in den nächsten Jahren die Sprachausgabe zu optimieren. Die kürzlich erfolgte Übernahme von Semantic Machines durch Microsoft, einem Unternehmen, das Konversations-KI entwickelt, zeigt wie wichtig diese Partnerschaft mit Alexa für die Einführung einer sprachgesteuerten Produktivitätssuite sein könnte. Vor kurzem wurde Microsoft ein Patent erteilt. Mit diesem Patent für eine so genannte „stille Spracheingabe“ (siehe auch Kap. 11) können Benutzer Sprachbefehle mithilfe von Flüstern aktivieren. Gegenwärtig verwenden die meisten Verbraucher Sprache ausschließlich in ihren Häusern und in ihrem Auto, und sie tendieren dazu, die Verwendung von Sprachbefehlen in der Öffentlichkeit zu vermeiden. Diese Technologie wird in einer Reihe von Geräten eingesetzt, von Handys über Smartwatches bis hin zu Fernbedienungen. Das Patentdesign vermittelt in seiner jetzigen Form das Gefühl von James Bond, in eine Smartwatch zu flüstern. Neue Ansätze für die Sprachbiometrietechnologie interessieren Microsoft sehr. Das Unternehmen hat kürzlich ID R & D, ein Unternehmen für Sprachbiometrie, in sein IoT-&AI-Insider-Labs-Programm aufgenommen.

Natürlich hat man auch bei Facebook verstanden, wie wichtig natürlichsprachige Kommunikation und dazugehörige Interfaces sind. Das Unternehmen ist sehr spät in den Markt eingetreten und dieser Beginn war eher holprig. Zunächst hat sich das Unternehmen sehr auf den Facebook Messenger fokussiert. Für Anwender und Entwickler war lange nicht klar, ob das Unternehmen Chatbots im Messenger begrüßt oder als Konkurrenz zu eigenen Angeboten versteht. Vor einiger Zeit hat Facebook wit.ai gekauft, ein Unternehmen, das eine interessante NLP-Engine per API anbietet und angekündigt hat, dass NLP in Kürze in

der Messenger-App standardmäßig zur Verfügung stehen soll. Wie genau dies allerdings zu einem für alle Teilnehmer attraktiven Ökosystem werden kann, ist unklar. Erkennbar kämpft Facebook in diesem Feld auch mit dem mangelnden Vertrauen im Bereich Datenschutz. Wie bereits an anderer Stelle erwähnt, ist Vertrauen – neben der Software – vermutlich eine der wichtigsten Fragen im Feld von CUIs und Assistenten. In einer vom Softwareunternehmen Voicyfy [4] in Auftrag gegebene repräsentativen Studie unter 1038 amerikanischen Erwachsenen sagten zwei Drittel der Befragten, dass sie auf unterschiedliche Weise besorgt sind über die Risiken für ihre Privatsphäre (Consumer Perception of Smartspeaker Privacy Risk). Dabei gehörten die Befragten zu Nutzern sprachgesteuerter Geräte.

Facebooks Versuch, einen eigenen Smartspeaker ausgerechnet mit Kamera und Video-Funktionalität auf den Markt zu bringen, stieß auf eher schlechte Marktresonanz. Viele Kommentatoren sprachen explizit die Frage an, ob Nutzer sich allen Ernstes eine Videokamera ausgerechnet von Facebook ins Haus holen sollten. Interessanterweise war dieses Gerät gleichzeitig eine Partnerschaft mit Amazon. Das Facebook-Portal ist über eine Sprachschnittstelle mit Alexa verbunden, damit Benutzer auf deren Fähigkeiten zugreifen können.

Wenn man die Geschwindigkeit der Entwicklung und die Intensität richtig einschätzen möchte, die Unternehmen weltweit derzeit im Feld von CUIs und intelligenten Assistenten entwickeln, dann muss man sich insbesondere auch mit dem asiatischen Teil der Welt befassen. In Kap. 7 tun wir dies ausführlicher. Huawei beispielsweise hat neben den eigenen Smartphones seit längerem intelligente Lautsprecher im Angebot und arbeitet an eigener Sprachtechnologie. Während die Smartphones noch Google Android als Betriebssystem funktionieren und damit auch in gewissem Umfang mit Google und Amazon zusammenarbeiten, strebt das Unternehmen mit seiner Sprachtechnologie außerhalb des chinesischen Marktes eine Eigenständigkeit an. Schon bevor Huawei durch die amerikanischen Handelsbeschränkungen dazu gezwungen wurde, sein eigenes Betriebssystem weiterzuentwickeln, um damit von Android unabhängig zu werden, kündigte der CEO von Huawei, Richard Yu, mehrfach an, dass er plane, seine KI-Technologie auch für andere Märkte außerhalb von China zu entwickeln. Dr. Chenglu Wang, President of Software Engineering of Huawei Consumer ergänzte auf einem Event in London im Januar 2019: „Huawei HiAI bietet Partnern und Entwicklern Cloud-Device-Chip-KI-Funktionen, die es ihnen ermöglichen, schnellere, intelligenter und einfachere mobile Anwendungen zu entwickeln. Huawei hat ein offenes KI-Labor in London und sechs KI-Labore in Europa eingerichtet und gemeinsame Forschungszentren mit mehr als 10 Universitäten aufgebaut. Mit Investitionen in Computervision, natürliche Sprachverarbeitung und andere Grundlagenforschung setzt sich Huawei für eine bahnbrechende Forschung und Anwendung von KI ein“ [5].

Wenn diese Erweiterung stattfindet, steht Huawei in direkter Konkurrenz zu Google, Apple und Amazon. Und genau das werden die anderen großen chinesischen Technologiemarken beobachten und bei ihrer Internationalisierung kopieren.

5.8 Marketingabteilungen als Temperaturfühler

Wer einschätzen möchte, wie wichtig ein Thema wird, unabhängig davon, ob es von den Technologieanbietern getrieben wird, wird bei Marketingverantwortlichen in Unternehmen einen Gradmesser finden. 88 Prozent der Marketingverantwortlichen (USA, Anfang 2019) glauben, dass Voice Assistants in den kommenden Jahren irgendwie wichtig sein werden. Immerhin ein Viertel glaubt, dass diese in zwei bis fünf Jahren ein außerordentlich bedeutender Marketingkanal werden [6]. Laut dieser Studie verstehen Marketers sehr gut, dass es dabei nicht nur um Smartspeaker, sondern auch um die Integration in Smartphones, Autos, TV-Geräten, Uhren und Wearables sowie alle Arten von Haushaltsgeräten geht. In der Umfrage unter 300 amerikanischen Marketingverantwortlichen wurde deutlich, dass Smartspeaker als Katalysatoren für die Sprach-Adaption bei Nutzern wirken. In der gleichen Studie wurde aufgezeigt, dass Smartspeaker-Besitzer zehn Prozent wahrscheinlicher auch bereits die Voice-Assistant-Funktionalität auf ihrem Smartphone benutzt haben, als diejenigen Smartphone-Besitzer, die noch kein Smartspeaker besitzen. Sie nutzen die Sprachfunktionalitäten auf dem Smartphone zu fast 40 Prozent täglich. So verstärken sich diese beiden Anwendungsbereiche wechselseitig.

So kommt es, dass nicht nur in den großen Unternehmen in Silicon Valley, Seattle und China Milliarden in Conversational User Interfaces und Assistenten-Technologien investiert werden, sondern auch in den Unternehmen an vielen Stellen Versuche unternommen werden – und hier die Investitionsaktivität täglich zunimmt.

5.9 Manches kann, vieles muss sich ändern

So richtig und notwendig es ist, dass Unternehmen beginnen, mit Alexa-Skills zu experimentieren und eigene Erfahrungen im automatisierten Dialog mit Kunden zu machen (siehe Kap. 2), so wichtig ist es auch, schnell zu verstehen, dass die Veränderungen stark in die Institutionen und Unternehmen hineinwirken werden und massive Auswirkungen auf Geschäftsmodelle, Informations- und Technologistrukturen der Unternehmen haben.

Der amerikanisch-indische Autor P.V. Kannan beschreibt diese fundamentale Veränderung in seinem 2019 erschienenen Buch [7]: „Das Zeitalter der Absicht: Einsatz künstlicher Intelligenz für ein hervorragendes Kundenerlebnis“ so: „Haben Sie sich jemals gewünscht, dass jedes Unternehmen, mit dem Sie interagiert haben, einfach weiß, was Sie wollen und es für Sie besorgen könnte? Dass, als Sie das Telefon abgenommen oder ein Chatfenster geöffnet haben, die Firma das, was sie über Sie wusste, nutzen würde, um Ihre Bedürfnisse vorwegzunehmen? Wir stehen kurz vor einer Zukunft, einfach so. Das Zeitalter der Absicht ist eine Welt, in der die intelligentesten Chatbots – virtuelle Assistenten – auf künstlicher Intelligenz basieren und mit der vollständigen Vergangenheit eines Kunden verbunden sind und den Kontext, in dem sich der Kunde befindet, einbeziehen. Diese virtuellen Assistenten können genau das vorhersehen, was

ein Kunde sucht, und Fragen im Chat, am Telefon und über intelligente Lautsprecher wie Amazon's Alexa beantworten. Sie werden die Geschäftswelt mit einem effizienten, skalierbaren Service verändern, der rund um die Uhr verfügbar ist und jeden Tag intelligenter wird.“ Wenn es technisch möglich ist, wird es in jedem Marktsegment einen Wettbewerber geben, der diese Art von „Hyper-Service“, der die Nutzerbedürfnisse in Echtzeit wahrhaft personalisiert und ohne Hierarchie und Qual der Wahl Entscheidungen erleichtert sowie lästige Vorgänge abnimmt, anbietet und alle anderen Anbieter plötzlich alt aussehen lassen.

5.10 Neue Regeln im Zeitalter der Absichten

Im Zeitalter der (erfüllten) Absicht, in der der Wunsch förmlich von den Augen abgelesen werden kann, gelten neue Regeln. Denn das Verstehen des Wunsches ist ja nur der erste Schritt. Um einen Wunsch auch passgenau erfüllen zu können, muss ich wissen und maschinenlesbar dokumentieren, wozu ein Produkt im Einzelnen überhaupt gut ist, denn nur so schließt sich der Kreis und eine Absicht kann erfüllt werden. Eine der neuen Regeln lautet, dass Marktintelligenz ohne Produktintelligenz keinen Nutzen mehr hat. In Produktkatalogen sind aktuell vor allem die Produktspezifikationen vermerkt, häufig aber nicht die spezifischen Anwendungsfälle oder gar explizit formulierte Fragestellungen, die potenzielle Nutzer vor sich haben, wenn sie nach einem Produkt suchen. Eine antihärtbeschichtete Pfanne wird im E-Commerce-Shop mit Größe, Material und Beschichtung vorgestellt. Ein Nutzer eines Einpersonenhaushalts wird künftig aber Fragen: „Alexa, ich suche eine Pfanne, bei der die Eier nicht anbrennen“ und wird erwarten, dass Alexa ohne weitere Fragen eine kleine teflonbeschichtete Pfanne schickt. Doch woher weiß Alexa, was Pfannen können, wenn Anbieter sie nicht „semantisch“, also in ihrem Kundennutzen beschreiben? Vermutlich wird deutlich, welche Veränderungen im Denken in Unternehmen ausgelöst wird, wenn diese Form von extremer Kundenzentrierung gefordert ist. Gleichzeitig wird deutlich, welche Mühe es kosten wird, alle Produktbeschreibungen auf dieses neue Format umzubauen.

Vergleichbare Herkulesaufgaben liegen vor den Unternehmen bei der Ertüchtigung ihrer IT-Infrastrukturen. Gespräche verlaufen in der Regel flüssig und schnell. Ein Wort gibt sprichwörtlich das andere und das ist es, was Nutzer zurecht von CUIs und intelligenten Assistenten erwarten. Schnelle Erkennung, notwendigenfalls ebenso schnelle Nachfragen und dann eine schnelle und passende Reaktion und Antwort. Nur wie kann dies gelingen, wenn Unternehmen ihre Daten nicht in Echtzeit auf Anforderung durch intelligente Assistenten liefern?

Ein Stottern entsteht, das man einem menschlichen Gegenüber richtigerweise verzeiht, aber nicht Siri und Google Assistant. Wie können Unternehmen ihre Schnittstellen und APIs so offen und zugänglich organisieren, dass Assistenzsysteme modular und bedarfsgerecht zugreifen können?

5.11 Die Veränderung ist massiver, als man zunächst denkt

Einen Eindruck über die Vielgestaltigkeit der kommenden Revolution bekommt man mit Blick auf das Verhältnis von Arzt und Patient. Conversational User Interfaces dringen in alle Bereiche von Wirtschaft und Gesellschaft. Ärzte haben eine Vielzahl von Dokumentationspflichten und sind daher im Patientengespräch häufig eher vor dem PC, als im Gespräch mit dem Patienten. Patienten und Ärzte leiden gleichermaßen darunter. Die Möglichkeit, diese notwendigen Aufzeichnungen durch Sprachsteuerung durchführen zu können, könnte die Begegnung zwischen Patient und Arzt vollkommen verändern und deutlich verbessern. Wie verändert sich das Gesundheitswesen, wenn Ärzte im Dialog die Vorgeschichte automatisiert präsentiert bekommen?

Das Beispiel zeigt, dass es sich nicht nur um ein „bisschen Sprachsteuerung für eigentlich das Gleiche“ handelt, sondern, dass sich die Aufgaben und Ausprägungen eines Berufes (praktische Ärzte) sowie den Nutzen, den seine Kunden (Patienten) erfahren, sich verändern. Natürlichsprachige Interfaces werden die Gesellschaft und Institutionen verändern, wie einst Autos das Stadtbild und die Architektur veränderten.

Neben den voraussichtlich in diesem Beispiel für die Patienten positiven Veränderungen, wird die Fähigkeit zur intelligenten automatisierten Transkription massiv in unser Berufsleben eingreifen und natürlich die Frage aufwerfen: „Wie sicher ist mein Job?“ Eine Frage, die in dieser sich anbahnenden Revolution gerade auch hoch qualifizierte Berufe betrifft.

Beispiel: Legal Tech. Schon lange verwenden Juristen Diktiergeräte um Schriftsätze und Sachverhalte zu erfassen. Maschinen, die das Gesprochene transkribieren sind ebenfalls schon länger verfügbar und bringen sehr gute Ergebnisse. Das Wortrepertoire ist wie bei Ärzten beschränkt und formelhaft. Somit können die Algorithmen effizient arbeiten. Neu ist dagegen, dass NLP auch Inhalte erfassen und Sinnzusammenhänge „verstehen“ kann.

Derzeit werden im juristischen Bereich vor allem Datenbanken, einfache Automationen, Such- und Filtertechniken eingesetzt. Aber es ist klar, wohin die Reise gehen wird: Juristische Suchdienste könnten NLP-Systeme einsetzen, um Suchanfragen tatsächlich zu „verstehen“ und das passende Dokument zu finden. In E-Discovery-Systemen könnten NLP- und Machine-Learning-(ML)-Technologien verwendet werden, um große Datensets inhaltlich zu verstehen und lediglich relevante Dokumente herauszufiltern. NLP- und ML-Tools könnten auch eingesetzt werden, um Verträge automatisch zu analysieren. Bei einer Vertragsprüfung könnten sie analysieren, ob ein Vertrag vollständig und in sich nicht widersprüchlich ist. Bei der Erfassung von umfangreichen Vertragswerken, beispielsweise beim Verkauf großer Gewerbeimmobilien, wird automatisch geprüft, ob alle Mietverträge vorliegen und welche Klauseln und Regelungen diese im Einzelnen enthalten. Eine weitere, technologisch anspruchsvollere Stufe wäre die Rechtsautomation, also die automatische Rechtssubsumtion bzw. Rechtsfindung. Programme könnten direkt eine juristisch valide Antwort für einen Sachverhalt präsentieren. Auch hieran wird bereits von einigen Firmen gearbeitet. An vielen Stellen sind Chatbots im Einsatz. Mit Hilfe von Frage-Antwort-Dialogen wird eine maßgeschneiderte juristische Beratung angeboten.

Durch die Fähigkeit, Wissensgebiete mithilfe von natürlichsprachigen Technologien zu bearbeiten, wird bei Ärzten und Anwälten nicht halt gemacht. Bei Banking und Financial Services, sowie im Verkauf und im Kundenservice wird sich die Welt derjenigen verändern, die sich bis vor Kurzem vor Automatisierung geschützt sahen.

Auf der Website Replaced by [Robot.info](#) [8] wird aufgezeigt, wie wahrscheinlich es ist, dass der eigene Beruf durch AI ersetzt wird. Der Wert für Radiologen liegt beispielsweise bei 100 Prozent. Auch wenn man natürlich davon ausgehen kann, dass es immer eine Anzahl von Berufsträgern geben wird, die die maschinellen Vorgänge überwachen und Auswertungen zur Entscheidung vorgelegt bekommen. Sei es als zweite Meinung, falls sich die Maschine nicht sicher ist oder nicht weiter kommt (beim Radiologen oder im Call Center eines Reiseanbieters). Dies gilt natürlich immer nur für einen Bruchteil der Menschen und ist auch nur gültig für diejenigen, die neue Werkzeuge annehmen und konstruktiv einsetzen.

Mehrere Start-ups arbeiten an Assistenten, die bei geschäftlichen Team-Meetings „anwesend“ sind und automatisiert ein Sitzungsprotokoll erstellen einschließlich der beschlossenen Aufgaben. Diese werden einzelnen Meeting-Teilnehmern zugewiesen und in deren Kalender eingetragen. Ein Assistent, der den (Arbeits-)Takt den Menschen vorgibt. Nicht mehr wie im Film bei Charlie Chaplin am Produktionsband, sondern in der Führungsetage.

Literatur

1. Harley Manning, Kerry Bodine, Josh Bernoff, Outside in: The Power of Putting Customers at the Center of Your Business, Brilliance Corp., 2014
2. Google: The Google Assistant – Investing in the future, <https://developers.google.com/actions/assistant-investments>, zugegriffen 23.11.2019
3. Nellis, Stephen: Apple, Salesforce team up to bring Siri to more business apps, (24.09.2018), <https://www.reuters.com/article/us-apple-salesforce/apple-salesforce-team-up-to-bring-siri-to-more-business-apps-idUSKCN1M41IM>, zugegriffen 23.11.2019
4. Voicify – Conversation Experience Platform, <https://voicify.com/>, zugegriffen 02.11.2019
5. Huawei: Technology Blending with Humanity: Huawei Ai makes the impossible possible, (25.02.2019), <https://consumer.huawei.com/en/press/news/2019/huawei-ai-makes-impossible-possible/>, zugegriffen 01.11.2019
6. Voicebot.ai: GIVING VOICE TO A REVOLUTION – THE STATE OF VOICE ASSISTANT AS MARKETING CHANNEL (2019), https://voicebot.ai/wp-content/uploads/2019/07/the_state_of_voice_assistants_as_a_marketing_channel_2019_voicebot.pdf, zugegriffen 20.11.2019
7. Kannan P.V.: „The Age of Intent: Using Artificial Intelligence to Deliver a Superior Customer Experience“, Amplify (2019)

Weiterführende Literatur

8. Will „Medical Transcription“ be Replaces bei Robots?, <https://www.replacedbyrobot.info/65026/medical-transcription>, zugegriffen 22.11.2019



Das sind die Player

6

Zusammenfassung

Wer einen CUI oder Assistenten für seine Marke entwickeln möchte, ist auf Hilfe angewiesen. In diesem Kapitel stellen wir die wichtigsten Player im Markt vor. Das reicht von CUI-Kanälen über Software-Integration bis zu Backend-Lösungen. Im Markt haben sich diverse Dienstleister für CUIs bzw. CUI-Apps positioniert. Falls Sie den Assistenten in Eigenregie entwickeln, ist es hilfreich, Ihren Entwurf bzw. Prototypen zu testen, um ihn vor dem Marktstart zu optimieren.

Mit dem neuen Feld von CUIs und intelligenten Assistenten entwickeln sich auch neue Konstellationen der Wertschöpfungsketten. Daher ist es interessant zu beobachten, welche Unternehmen und anderen Akteure, neben den Anwendern von Conversational User Interfaces und intelligenten Assistenten, sich im entwickelnden Ökosystem bewegen.

CUIs und intelligente Assistenten entwickeln sich in verschiedenen Regionen der Welt unterschiedlich in Geschwindigkeit und Qualität. Daher sind Marktteilnehmer in Europa, USA, China und Indien nicht immer deckungsgleich. In diesem Kapitel beleuchten wir vorrangig den westlichen Markt und versuchen aufzuzeigen, welche Arten von Spielern sich derzeit herausbilden. Im nächsten Kapitel beschäftigen wir uns vor allem mit den Konstellationen im chinesischen und indischen Markt.

Alle Felder der Wertschöpfungskette decken die großen Plattformbetreiber (z. B. Google) ab. Sie betreiben eigene natürlichsprachige Assistenten und sprechen ihre Endnutzer über ihre eigenen Kanäle direkt an. Darüber hinaus bieten Sie eigene Clients und Hardware an, über die ihre eigenen Assistenten genutzt werden können. Daneben gibt es Anbieter, die viele oder alle notwendigen Module zum Betrieb von CUIs als Cloud-Dienstleistungen anbieten (z. B. IBM). Teilweise agieren die großen Plattformbetreiber

allerdings auch noch als Anbieter modularer Cloud-Services. Dienstleister und Agenturen bedienen sich dieser Cloud-Dienste, kombinieren diese ggf. mit eigenen Softwarelösungen und erledigen die Integration in die Backendsysteme der Unternehmen, die ihre eigenen Plattformlösungen schaffen, die über Kanäle und Clients der großen Anbieter, oder auf ihren eigenen Websites und Mobile-Apps, an die Endnutzer herantragen. Immer mehr Anbieter von Hardware (z. B. Küchengeräten, Lautsprecher, Autos) integrieren die Assistenten der großen Plattformanbieter in Ihre Geräte.

Die großen Plattformunternehmen sind nicht nur in einem der erwähnten Felder tätig sind, sondern in mehreren gleichzeitig. So bietet Google mit dem Google Assistant eine eigene Plattform, hat mit dem Google Assistant einen eigenen Kanal, bietet mit Nesthub und Smartphones eigene Clients und bietet mit Dialogflow eine Cloudlösung für Drittanbieter an. Mit dieser breiten Abdeckung der gesamten Wertschöpfungskette (siehe Abb. 6.1) erscheint Google wie ein eigenes kleines Ökosystem.

6.1 Die Plattformen, CUI-Kanäle und -Endgeräte

Weltweit gibt es derzeit ungefähr 13 große Universal-Plattformen, die aus den USA oder China kommen und in ihren Heimatmärkten und darüber hinaus angeboten werden.

Google im Westen und das chinesische Tencent mit seinem Angebot WeChat sind Beispiele für derartige Universal-Plattformen. Amazon, Apple, sowie deren chinesische Pendants, Baidu und Alibaba stellen weitere Beispiele dar.

Alle diese Unternehmen bieten Plattformen, die sich an Endnutzer wenden, in denen Conversational User Interfaces Zugang zu intelligenten Assistenten geben. Meistens nicht beschränkt auf Voice, sondern in unterschiedlichen Ausprägungen – auch in Form von

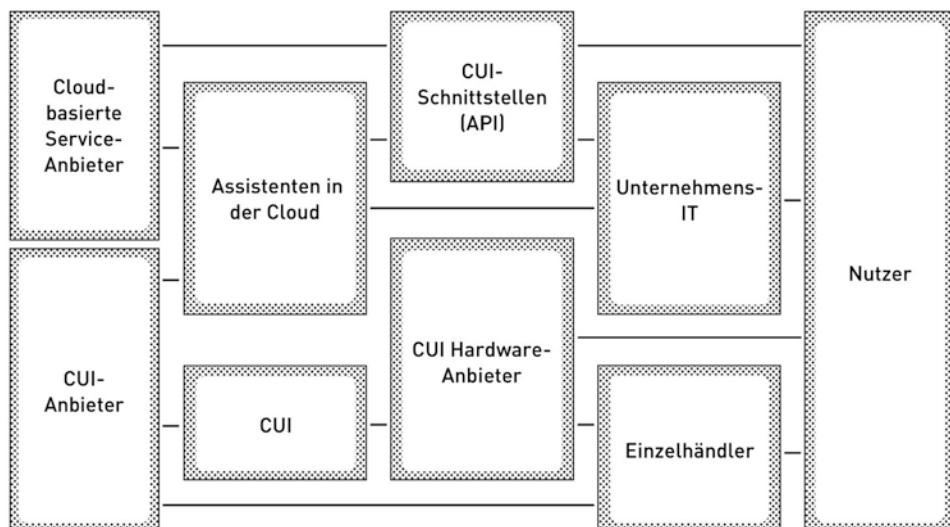


Abb. 6.1 Wertschöpfungskette von CUI und Assistenten

Texteingabe. In diesen Komplettangeboten sind jeweils die notwendigen Fähigkeiten zur Spracherkennung enthalten (NLP), die Universal-Plattformen haben darüber hinaus unterschiedliche Basis-Service-Integrationen und aggregierten Inhalte – Amazon Alexa bietet neben dem Zugriff auf den Amazon-Shopping-Katalog auch Antworten auf Wetterfragen und zu allgemeinen Themen, sowie Dienste wie Timer etc. Drittanbieter können auf verschiedene Weise App-ähnliche Programme erstellen und damit eigene Angebote machen. Bei Amazon heißen diese Programme Skills, was ihre Funktion sehr gut beschreibt, denn diese Programme, die teilweise auch Voice-Apps oder Bots genannt werden, haben meistens einen beschränkten Funktionsumfang. Bei Google heißen sie Actions, bei Samsung Capsules. Die Plattformen tragen klangvolle Namen wie Alexa (Amazon), Siri (Apple) TMall Genie (Alibaba).

In China ist WeChat die vermutlich am weitesten entwickelte Plattform. WhatsApp, das häufig als westliches WeChat wahrgenommen wird, hat die Entwicklung zu einem vollintegrierten Anbieter bislang nicht vollzogen und dadurch vielleicht den Anschluss verpasst. Gleichzeitig gibt es Messenger-Plattformen in gewissen Marktsegmenten, wie Telegram, die besonders in einigen asiatischen Ländern und dem Nahen Osten ihre Verbreitung haben. Sie bieten zwar ebenfalls keine Integration von Funktionen wie NLP und Payment an, verfügen allerdings aber deutlich offenere Schnittstellen für Drittanbieter, die darauf aufbauend Conversational Systeme und funktionale Bots integrieren.

Eine besondere Bedeutung kommt naturgemäß denjenigen Anbietern mit großer Nutzerbasis und eigener Hardware-Plattform zu. Apple und Google, die neben ihren Assistenten-Services auch eine breite Basis von Smartphonenuutzern und mit iOS und Android ein eigenes Betriebssystem haben, können in der Integration von Hard- und Software nicht nur sehr schnell leistungsfähige Systeme anbieten, sondern durch den vollen Zugriff auf die Nutzungsdaten diese Systeme auch ebenso schnell optimieren. Diese Sonderstellung wird künftig noch stärkere Bedeutung erhalten. Das wird deutlich, wenn bestimmte Funktionsbestandteile – wie beispielsweise die Sprachverarbeitung – direkt auf den in Smartphones oder Speakern eingebauten Chips stattfinden und dadurch um ein Vielfaches schneller werden, als dies mit gängigen Cloud-Lösungen möglich ist. Auch die Fragen rund um den Schutz der Privatsphäre können besser beantwortet werden (siehe dazu auch Abschn. 5.6). In der Regel werden heute die Sprachaufnahmen zur Analyse über eine mobile Datenverbindung in ein Rechenzentrum geschickt. Hier erfolgt die Auswertung per NLP und die Antworten gelangen auf gleichem Weg zurück zum Nutzer. Das Ergebnis sind längere Reaktionszeiten, auch Latenzen genannt.

Hardware-Anbieter wie Samsung, Huawei und in etwas anderer Form auch Xiaomi versuchen, den intelligenten Assistenten in die von Ihnen angebotenen Smartphones und anderen Geräte zu integrieren und auf diese Weise Funktionsvorteile und Nutzerakzeptanz zu erzielen. Die Breite der Produktpalette von Samsung – Smartphone, Fernseher, Kühlschränke, Geschirrspülmaschinen, Waschmaschinen und Öfen – bietet hier ungewöhnliche Möglichkeiten. Diese Chance scheint Samsung ebenfalls zu erkennen. Samsung verfolgt die Strategie eines eigenen CUI-App-Marktplatzes [1], der alle Schnittstellen zwischen der Vielzahl von Geräten und Apps/Anwendungen von Drittanbietern miteinander vernetzt und kompatibel machen soll (Siehe Kap. 9). Auf diese Weise können

TV-Geräte mit Küchengeräten, Smartspeakern und Smartphones mühelos mit dem Samsung Assistenten Bixby gesteuert werden. Eine gute und richtige Idee, die allerdings derzeit noch nicht genügend Entwickler begeistert, weshalb bislang nur wenige attraktive Anwendungen entwickelt wurden. Das gilt auch für Huawei, dessen HiAssistant im chinesischen Markt naturgemäß weiter verbreitet ist als in westlichen Märkten. Doch im Heimatmarkt kann sich der Huawei Assistant nicht gegen die Wettbewerber durchsetzen. Im Fall beider Unternehmen zeigt sich, wie schwer es ist, ohne eigenes Smartphone-Betriebssystem – beide Anbieter lizenzierten Android – seinen Assistenten optimal auf das Smartphone abzustimmen. Xiaomi bietet im chinesischen Markt eine zunehmende Anzahl von Home-Automation-Lösungen und -Gadgets, darunter Überwachungskameras und Türschlösser, die aus der Ferne geöffnet werden können. Diese Geräte werden allesamt über den Xiao genannten Sprachassistenten von Xiaomi gesteuert.

Amazon hat durch seinen frühen und beherzten Markteintritt mit Amazon Echo Smartspeakern, auf denen der Assistent Alexa läuft, eine sehr gute Marktstellung erreicht und ist unter den Plattformbetreibern international auf einer Führungsposition. Amazon hat damit, zumindest in Europa, etwas erreicht, was sehr selten gelingt. Alexa ist zum Gattungsbegriff für Smartspeaker und intelligente Assistenten im Allgemeinen geworden. Allerdings fehlt Amazon derzeit eine weitergehende Integration in ein eigenes Smartphone-Betriebssystem, die Apple und insbesondere Google derzeit so stark macht. Die Anzahl und Vielzahl der Endgeräte, mithilfe derer Nutzer auf CUIs und Assistenten zugreifen können, nimmt täglich zu. Neben den großen Drei: Smartspeaker und Smartphone und Auto, kommen täglich neue Devices hinzu: Kopfhörer mit unterschiedlichen Funktionen, über die Assistenten angesprochen werden können; Smartwatches und andere Wearables werden mit Mikrofonen ausgestattet und dienen als Zugang zu Voice-Assistenten. Apple, Google und vermehrt auch Amazon bringen Produkte in all diesen Kategorien auf den Markt. Zunehmend dienen auch Laptops, TV-Geräte und Game-Konsolen als Zugangsgeräte zu CUIs und Haushaltsgeräten aller Art. Alle diese Geräte haben – teilweise unsichtbare – Clients mithilfe derer CUIS und Assistenten genutzt werden können.

6.2 CUI-Plattformen – Software-Integratoren und die Allrounder

Betrachtet man den Markt, gibt es neben Anbietern mit eigener Softwareplattform und Endnutzern auch starke Anbieter, die dem Nutzer vermutlich zunächst nicht viel bedeuten und dennoch im Markt eine wichtige Rolle spielen: Gemeint sind Allrounder wie IBM, Microsoft, Oracle, SAP und Adobe. Sie bieten Full-Stack Cloud-Services an, mithilfe denen Unternehmen eigene Conversational Systeme aufbauen und betreiben können. Diese Cloud-Services umfassen verschiedene Elemente von Natural Language Processing, aber auch andere so genannte Cognitive Services. Google ist ebenfalls ein Allrounder, denn das Unternehmen bietet Grundlagendienste wie NLP, Hosting und AI aller Art unter Google API.ai. Diese Dienste können von Unternehmen zum Aufbau und Betrieb

eigener und eigenständiger CUI-Anwendungen eingesetzt werden, ohne dabei automatisch im Google-Endkunden-Dienst Messenger integriert zu sein. IBM bündelt seine Dienste unter der Marke IBM Blue Mix, wo eine Vielzahl von Services mittels APIs angeboten werden, die auf der künstlichen Intelligenz von Watson aufbauen. Gleichzeitig versucht IBM Watson – mit wenig Erfolg – auch als eigenständige Endkunden-Assistenten-Marke anzubieten. Auch Microsoft ist früh ins Thema der Assistenten und Sprachsteuerung eingestiegen und hat in seine Softwaresuite Office den eigenen Sprachassistenten Cortana eingebaut. Allerdings scheint Microsoft es aufgegeben zu haben, unabhängig von Cortana vom eigenen Desktop-Betriebssystem an Endkunden zu vermarkten. Sie betreiben stattdessen mit LUIS.ai und MSG.ai vergleichbare Angebote und bieten mit dem Microsoft-Bot-Framework eine spezielle Cloud-Software-Suite, um Chatbots zu erstellen. Gleichzeitig bietet Microsoft eine ganze Reihe von kognitiven Services in der Cloud an, die Betreiber von Assistenten buchen können, und es öffnet Cortana in verschiedene Richtungen: Einerseits indem interessierte Unternehmen Bestandteile von Cortana-Fähigkeiten in ihre eigenen Angebote integrieren können, ohne dass der Endnutzer dies bemerkt, aber auch, indem Cortana als App/Skill beispielsweise in Amazons Alexa angeboten wird. Selbst das sonst eher verschlossene Facebook hat mit Wit.ai eine offene Schnittstelle für dialogische Systeme und unterstützt damit Unternehmen darin, unabhängig von Facebook, eigene Assistenten zu erstellen.

Diese großen Allrounder bieten allerdings auch speziell auf bestimmte Branchen, oder Unternehmensgrößen ausgerichtete Anwendungslösungen an, in denen die oben genannten Einzeldienste gebündelt sind. [CallJoy.com](#) (siehe Kap. 10) beispielsweise ist ein Angebot für kleinere und mittlere Unternehmen, mit dem diese eingehende Kundenanfragen per SMS, Messenger-Dienst oder Sprachanruf automatisiert bearbeitet und damit das faszinierende Projekt Google Duplex in der Praxis einsetzbar macht.

6.3 CUI-Backends und Businesslösungen

Neben Anbietern mit international bekannten Namen, gibt es ein wachsendes Angebot von weniger prominenten, gleichwohl sehr leistungsfähigen Lösungen, wie beispielsweise der schwedisch-spanischen [Artificial-Solutions.com](#). Sie bieten dialogische Systeme für Unternehmen an, die mit leistungsfähigen NLP und Dialogkomponenten Dialoge über eine Vielzahl von Eingabe- und Ausgabekanälen betreiben kann, so dass Kunden mit dem Unternehmen am Vormittag über E-Mail, während der Mittagspause über Facebook Messenger und am Abend im Chat auf der Unternehmenswebsite ihre Anliegen weitgehend automatisiert im Dialog bearbeiten können.

Das englisch-indische Unternehmen Gupshup.io bietet Vergleichbares und öffnet sich noch stärker als Selfservice-Lösung. Es ermöglicht Unternehmen, eigenständig derartige Systeme aufzubauen und zu betreiben. Ein Ansatz, den weltweit hunderte von anderen Anbietern verfolgen. Nicht wenige dieser Anbieter kommen aus Indien, wo sich SAAS-Angebote (Software as a Service) und auch NLP stärker entwickelt haben.

Unternehmen dabei zu unterstützen, ihre Kundenkontakte zu verbessern und möglichst weitgehend zu automatisieren, entwickelt sich international zu einem Wachstumsmarkt. Bereits früh in den 1970er- und 1980er-Jahren haben Anbieter von Großtelefonanlagen einfache Möglichkeiten zur Kontaktautomatisierung angeboten. Dabei handelte es sich um wenig flexible Hard- und Softwarekombinationen, deren primitive Programmierung in wenig transparenten Softwareumgebungen nur von den Anbietern selbst vorgenommen werden konnte. Um die Jahrtausendwende entstanden Unternehmen, wie Nuance in den USA oder Novomind aus Hamburg, die gewissermaßen in einer nächsten Generation von der Telefonanlagen-Hardware unabhängige Softwarelösungen entwickelten, dabei allerdings meist proprietäre Standards einsetzen. Diese Anbieter entwickeln sich nun schnell weiter, verlagern ihre Anwendungen teilweise in die Cloud und nutzen Schnittstellen mithilfe derer externe Dienste, wie beispielsweise die biometrische Auswertung von Stimmen, eingebunden werden können. Das an der Nasdaq notierte Unternehmen Liveperson gehört ebenfalls in diese Anbieterkategorie und hat durch Zukäufe eine umfangreiche Software-Suite aufgebaut. In diesem Zusammenhang sollte auch auf die große Zahl von Unternehmen verwiesen werden, die sich auf einzelne Aufgabenfelder konzentrieren. Beispielsweise die 2009 gegründete [GoVivace.com](#), die sich auf Speechrecognition, Voice-Biometrics, Speaker-Identification und Gender-Identification spezialisiert hat. Ihre Softwarelösungen können in bestehende Firmen-Technologie-Architekturen eingebunden werden.

Im Bereich der Ausgabe gibt es eine ganze Reihe interessanter Softwareanbieter, die automatisiert Antworten und Texte erstellen, die auf teilweise komplexe Nutzeranfragen aus Datenbanken on-the-fly antworten. Neben automatisiert erstellten Börsenberichten, die basierend auf den aktuellen Kursentwicklungen News-Flashes produzieren, gibt es automatisiert erstellte Produktbeschreibungen. Darüber hinaus finden sich in dem Bereich etwas überraschende Anwendungsfelder: Das Stuttgarter Unternehmen AX-Semantics entwickelt Algorithmen, die in der produzierenden Industrie von Robotern und Großanlagen eingesetzt werden, um Sprachdialoge mit den Arbeitern zu ermöglichen. Das passiert nicht nur auf Deutsch, sondern in unterschiedlichen Sprachen, je nach Herkunft des Maschinenführers.

Mit welcher Stimme ein Assistent spricht, ist ebenfalls Gegenstand unterschiedlicher B2B-Angebote. Sofern das Angebot auf einer der Universalplattformen aufbaut, kann man die synthetische Stimme von Alexa oder Siri einsetzen. Häufig suchen Unternehmen nach Möglichkeiten, sich von diesen als wenig spezifisch wahrgenommenen Stimmen zu lösen und eigene Angebote zu machen. Dabei reicht die Bandbreite der Möglichkeiten von eigenen synthetischen Stimmen, bis zu so genannten Voice Actors. Es gibt verschiedene Anbieter, die Voice-Synthese-Software anbieten, die unter Umständen klingen kann, wie die Stimme einer bekannten Persönlichkeit, sei es der CEO des Unternehmens oder ein bekannter Schauspieler. So könnte ein Nespresso-Bot klingen wie George Clooney. Immer häufiger werden auch Voice Actors eingesetzt. Echte Sprecher, die in der Regel gescriptete Antworttexte von Assistenten in mühevoller Arbeit mit hunderten Variationen einsprechen. Auch hierfür gibt es Anbieter, die die richtigen Werkzeuge zur Erstellung derartiger Sprachbibliotheken anbieten.

6.4 Open Source und On-Premisse

Auch im Feld der CUIs und intelligenten Assistenten gibt es Open-Source-Initiativen. Aus deutscher Sicht erfreulich ist ein Berliner Unternehmen, das weltweit eingesetzt wird und mit Rasa.ai eine Open-Source-Bibliothek anbietet. Sie kann zur Intent-Erkennung, Entity-Extraktion und verschiedenen Aspekten von Dialog-Management eingesetzt werden. Anders als die meisten in diesem Kapitel erwähnten Angebote der kommerziellen Unternehmen, können Firmen mit Rasa.ai CUIs und intelligente Assistenten aufbauen, die vollständig lokal, auf den Servern des Anbieters, gehostet werden. Damit bieten sie eine brauchbare Lösung für sensible Daten, die Unternehmen aus verständlichen Gründen nicht auf Servern von Google oder Amazon und mit Verarbeitungssoftware von Microsoft oder IBM bearbeiten lassen wollen.

Die Sensibilität von Daten motivierte auch Rand Hindi [2], den Gründer von Snips, einem Smartspeaker, der als Open-Source-Lösung konzipiert ist und dessen Sprachverarbeitung vollständig lokal erfolgt. Damit haben Google, Amazon usw. keinen Zugang zu persönlichen Daten. Auch, wenn es möglicherweise eher zweifelhaft ist, ob Snips nachhaltigen Erfolg haben wird, zeigt es doch, dass die Idee von offenen Plattformen, die dem aktuell eher verbreiteten Konzept von „Walled Gardens“ und dem damit verbundenen zunehmenden Lock-In-Effekt von Anbietern und Nutzern entgegenstehen, von vielen kritisch gesehen wird.

6.5 Die CUI-App-Anbieter

Alle Plattformanbieter ermöglichen Drittanbietern, Inhalte und Dienste über Ihre Kanäle und Clients zu veröffentlichen. Die Regeln sind dabei sehr unterschiedlich und werden, wie bei jeder Form von geschlossenen Betriebssystemen, weitgehend von den Plattformanbietern definiert. Das System ist prinzipiell geschlossen – ein sogenannter Walled Garden – und damit eher mit der aus Mobile bekannten Struktur von Betriebssystem und Apps vergleichbar, als mit der des offenen und in gewisser Weise freien Internets. Aufgrund der schnellen technischen Weiterentwicklung von wichtigen Komponenten ändern sich die Regeln und Möglichkeiten der Gastgeber-Plattformen derzeit schnell. Ähnlich wie bei Apps entstehen Wellen, in denen Anbieter von anderen digitalen Angeboten ihre auf Websites und Apps verfügbaren Inhalte und Dienste mehr oder weniger gut in das neue Umfeld umsetzen und mit Actions und Skills experimentieren und damit teilweise bereits gute Ergebnisse erzielen. In der Regel beginnen Unternehmen zunächst damit, isolierte Projekte zu betreiben und diese mit der nächsten Iteration mit eigenen Prozessen zu verzahnen und beispielsweise den Kundenservice oder die Produktauswahl mittels CUI und intelligenten Assistenten zu verbessern. Neben diesen Anbietern, die bereits Mobile-Apps und Websites betreiben und deren Funktionen für CUIs übersetzen, kommen mit dem neuen Feld auch neue Pure-Player auf den Markt, die Voice-Only arbeiten und nicht selten auf diese Weise für Nutzer interessante Angebote auf den Markt bringen und schon gleich zu Beginn der Marktentwicklung zu ernsten Wettbewerbern werden.

6.6 Die Dienstleister

Natürlich entsteht nicht nur Nachfrage und Angebot im Feld der eigentlichen Software, sondern auch ein wachsender Bedarf an Dienstleistern, die potenzielle Anbieter von CUIs und intelligenten Assistenten bei Planung, Umsetzung und Betrieb ihrer Chatbots und Voice-Assistenten unterstützen. Neben den großen Unternehmensberatern und Systemintegratoren gibt es eine wachsende Zahl von kleineren, unabhängigen Unternehmen, die in dem jungen Markt teilweise herausragende Arbeit leisten. Häufig entwickeln diese Beratungsunternehmen eigene kleinere Softwareframeworks, die zunächst nur für den internen Gebrauch gedacht sind und sich dabei in der täglichen Arbeit als derart nützlich und leistungsfähig erweisen, dass die Unternehmen dann beginnen, daraus Softwareprodukte zu formen, die eigenständig, oder zumindest teilweise eigenständig vermarktet werden.

Ein gutes Beispiel dafür ist das Unternehmen Mercury.ai, das aus Bielefeld stammt und ein Spinoff der dortigen Hochschule ist, die als eines der Excellence-Cluster für Sprache und Machine Learning im deutschsprachigen Raum gilt. Begonnen hat Mercury.ai als Agentur, die große Konsumgüteranbieter beim Entwickeln von leistungsfähigen Chatbots half. Das in diesem Zusammenhang entwickelte Bot-Framework wuchs stetig und heute bietet es Mercury.ai unabhängig als Werkzeug an. Mithilfe dessen kann der Prozess vom Bot-Prototyp zum betriebsfähigen Multi-Plattform-Bot gut bearbeitet werden.

Einen vergleichbaren Weg geht das Berliner Unternehmen Future of [Voice.com](#), welches ebenfalls als Agentur begann. Es unterstützt Unternehmen bei Planung und Aufbau von Alexa Skills. Außerdem entwickelt es derzeit eine Art Content-Management-System für Voice-Applikationen.

Natürlich gibt es auch Agenturen, die sich ausschließlich auf Konzeptentwicklung und Umsetzung im Kundenauftrag konzentrieren. Rain.agency aus New York betreibt Büros in weiteren US-amerikanischen Städten. Es macht durch kreative Auftragsarbeiten für große amerikanische Marken von sich reden, darunter für den Sportartikelanbieter Nike, die Football-League NFL und das Waschmittel Tide.

Wer sich auf den amerikanischen Fachkongressen umsieht, trifft Monat für Monat auf mehr kleine und größere Spezialagenturen, die von Bot-Strategie bis zur Umsetzung, laufender Betreuung der Bots bis zu Voice-Anwendungen viele Dienstleistungen anbieten. Teilweise fühlt man sich dabei an die stürmische Entwicklung des Internets und der Spezialagenturen Mitte der 1990er-Jahre oder an die vergleichbare Zeit Mitte der 2000er-Jahre rund um das Thema Mobile erinnert.

[Grandstudio.com](#) aus Chicago ist eine User-Experience-Agentur, die dadurch hervorsticht, dass sie Voice Interfaces auch im Kontext von Inklusion und Barrierefreiheit betrachtet und somit eine Reihe von sozialen Use Cases bearbeitet.

Im deutschsprachigen Raum haben sich bislang nur wenige auf Conversational User Interfaces spezialisierte Agenturen entwickelt. Häufig bieten Digitalagenturen das Thema mit an. Andere Dienstleister kommen eher aus dem Technologiebereich und erweitern ihre Dienstleistungen um inhaltlich-konzeptionelle und kreative Aspekte.

6.7 Testen, Optimieren und Prototypen

Ein bekanntes Sprichwort aus der Szene der Interaction Designer besagt: „Was man nicht testen kann, kann man auch nicht verbessern“. Daher bleibt ein Überblick über die Akteure im wachsenden Markt unvollständig, ohne einen Blick auf das Feld der Analyse- und Monitoring-Software sowie ihrer Anbieter.

Es war bereits die Rede von der Schwierigkeiten, CUIs und intelligente Assistenten zu testen. Anders als beim Test von gewöhnlichen Websites und Apps geht es ja nicht nur um möglichst fehlerfreie Funktion, sondern darum, vernünftige und nützliche Ergebnisse für Nutzeranfragen zu evaluieren. Die Antwort auf die Frage, wie nützlich und sinnvoll die Antwort von Alexa oder einem individuellen Sprachassistenten tatsächlich ist, hängt naturngemäß stark vom einzelnen Nutzer und seinem individuellen Kontext ab. Außerdem gibt es bei CUIs keine festen Bildschirme, die man für Testcases präparieren kann. Es bleibt nur der Dialog, sei es über Text oder Sprache. Dieser Dialog ist bei jedem Nutzungs-vorgang individuell und vollständig personalisiert. Das US-amerikanische Unternehmen [Cyara.com](#) entwickelte ursprünglich Testsysteme für IVR-Systeme. Nun bietet es auch „intelligente Testsysteme“. Die Testroutinen entsprechen „echten Nutzer-Journeys“, die zunächst aufgezeichnet wurden, automatisierte Varianten entwickeln, beispielsweise andere Adressdaten einfügen, automatisiert verwandte Fragestellungen aufrufen und damit den Bot in vielen verschiedenen Dimensionen testen.

Einen etwas anderen Ansatz betreibt [Cognilytica.com](#) ein Beratungsunternehmen, das strukturierte Evaluationen von AI-Software aller Art betreibt und die Kunden anschließend dabei unterstützt, ihre jeweiligen Services zu optimieren. Cognilytica erstellt in regelmäßigen Abständen einen Vergleich der großen Conversational Services (Google, Apple, Amazon etc.) und vergleicht die Services seiner Kunden gegen diese Panels.

Auch im Feld der Test- und Evaluationsanbieter gibt es erfreulicherweise europäische Anbieter. So stammt das Unternehmen Botium.at, das eine leistungsfähige Software zum Test von Chatbots aller Art anbietet, aus Österreich.

Ähnlich bedeutend wie die Analyse von Bots und Sprachassistenten ist die Frage, wie diese konzipiert und entwickelt werden. Prototyping – also die zunächst bewusst beschränkte Funktion und Fähigkeit – um damit den tatsächlichen Funktionsumfang im Dialog mit anderen Teammitgliedern und ersten Testnutzern definieren zu können. Prototyping ist im Feld von Websites und Apps seit Jahren gebräuchlich und entsprechend gibt es eine große Zahl von Tools großer und kleinerer Anbieter, die die Erstellung von Websites und Apps auch für Designer ohne das Hinzuziehen von Softwareentwicklern ermöglicht. Die mit diesen Tools entwickelten Dummies sind aus den Produktionsprozessen nicht mehr wegzudenken. Viele der zum Einsatz kommenden Prototyping-Werkzeuge sind so ausgelegt, dass einzelne Elemente, beispielsweise Grafiken, im späteren „echten“ Entwicklungsprozess ohne großen Aufwand wieder verwendet werden können. Im Feld von CUIs und intelligenten Assistenten ist die Anzahl brauchbarer Prototyping-Werkzeuge und -Anbieter noch sehr eingeschränkt: Bei Botsociety.io und [BotMock.com](#) handelt es

sich um Startups aus den USA, die in diesem Feld die ersten Produkte auf den Markt gebracht haben. Damit haben sie in gewisser Weise die Art definiert, wie CUI-Prototypen aussehen und funktionieren. Das Schwergewicht Adobe hat den Markt ebenfalls erkannt und ist dabei, sein im Bereich von Websites und Apps gut eingeführtes Prototyping-Werkzeug XD [3] durch entsprechende Funktionen für den Einsatz bei CUIS und intelligenten Assistenten zu ertüchtigen. Es bleibt spannend zu beobachten, ob damit der Wettbewerb und das Ringen um die beste Lösung in diesem wichtigen Markt bereits abgeschlossen ist – oder ob es noch vollkommen neue Ansätze geben wird.

6.8 Lösungen in Kombination aus Hard- und Software

CUIs und intelligente Assistenten benötigen Geräte und Hardware auf denen diese installiert sind und vom Anwender genutzt werden. Zunächst denkt man dabei an Smartspeaker, die gewissermaßen die sichtbarsten Inkarnationen von CUIs sind. Wir alle haben ein Bild von den Alexa- und Google-Home-Geräten. CUIs und intelligente Assistenten sind aber bereits heute in einer großen Zahl anderer Gerätetypen verfügbar. Die Anzahl und Vielfalt wird sich künftig noch erheblich steigern.

Neben Smartspeakern sind Smartphones vermutlich die Hardwareplattform, auf der die größte Zahl integriert ist. Betrachtet man die Verbreitung von Android und iOS, wird dies voraussichtlich noch eine ganze Zeit so bleiben, selbst wenn die Verbreitung von Smartspeakern weiterhin so stark wachsen wird, wie dies aktuell der Fall ist.

Die Hardwareplattform Smartphone ist nicht nur quantitativ interessant, sondern auch qualitativ. Das hat verschiedene Gründe: Einerseits ist das Smartphone buchstäblich immer dabei. Es wird zu Zeiten und an Orten genutzt, an denen Smartspeaker vermutlich noch lange Zeit nicht gelangen werden. Außerdem sind Smartphones mit den eingebauten Touchscreens die natürliche Basis für ein multimodales Nutzererlebnis. Sie sind für gesprochene Stimme, getippten Text, Bilder und Sounds und jedes weitere heute denkbare Nutzungsszenario optimal ausgestattet. Dazu zählt auch der Umstand, dass Smartphones ganz natürlich einen Strom von Nutzerdaten generieren. Wie wir bereits dargestellt haben, funktionieren alle wesentlichen Elemente von CUIs und intelligenten Assistenten besonders gut, wenn viele Informationen über Kontext, also Wo, Wann, Was, Warum, Was war zuvor, dem Assistenten zur Verfügung stehen und in den Prozess eingebaut werden können. Zuletzt ist das Smartphone ein Hochleistungscomputer, auf dem komplexe Algorithmen des NLP problemlos und schnell durchgeführt werden können. Im Zusammenhang mit dem üblicherweise vorhandenen Speicher auf den aktuellen Geräten können Spracherkennung, Authentifizierung und andere kognitive Dienste in beeindruckender Weise angeboten werden. In diesem Sinne sind Anbieter, die Smartphones mit eigenem Betriebssystem auf dem Markt haben – derzeit ausschließlich Apple und Google – in einer herausragenden Position. Sie sind den übrigen Playern (Amazon, Xaomi, Alibaba) auf absehbare Zeit eindeutig überlegen.

Anbieter von Geräten aller Art werden von den Plattformbetreibern derzeit umworben. Das Ziel: Ihre Assistenten sollen Kopfhörer und Lichtschalter, Glühbirnen und Mikrowellen steuern. Über Lizenzvereinbarungen sind die Assistenten längst in das Ökosystem der Smart-Home-Geräte eingebunden. Auf diese Weise kann man das TV-Gerät von Samsung mit Amazon und Google steuern und Musiksysteme von Bose oder Sonos kommen mit Google Assistant und Amazon Alexa inside. Es ist abzusehen, dass wir mit dem Kühlenschrank ins Gespräch kommen, das Licht anschalten und die Heizung regulieren. In jedem dieser Geräte haben wir mehrere intelligente Assistenten zur Auswahl. Sobald diese Geräte Bildschirme integrieren (z. B. Kühlchränke) werden diese über kurz oder lang in das multimodale Gesamtsystem eingebunden, so dass man die Endauswahl der Shopping-Assistenten-Shortlist auf dem Smartphone, dem TV-Bildschirm oder eben im Vorbeigehen auf dem Kühlchränkbildschirm erledigen kann.

Das wichtigste aller Geräte, die Assistenten-Träger sein können und einen Bildschirm haben, ist natürlich das Auto. Der Kampf um die Integration ins Auto war in den vergangenen zwei Jahren das Hauptfeld des Wettkampfes zwischen Google, Apple und auch Amazon. Alle drei versuchen, ihre Systeme besonders autotauglich zu machen und damit den Zugang zu ihren Nutzern in diesem international wichtigen Use Case auszubauen. Selbstverständlich wissen auch die Automobilhersteller um den Wert der Schnittstelle im Auto zu ihren Kunden. Doch ist ihnen die Kontrolle über diesen Kundenzugang bereits bei Einführung der Mobiltelefone verloren gegangen. Fahrer und Nutzer verlangen selbstverständlich ihre Telefonbücher auf dem Handy im Auto nutzen zu können und die Playlists vom iPhone im Display zu sehen. Das Verhältnis zwischen Autoherstellern und Anbietern von Mobil-Betriebssystemen ist kompliziert, immerhin lässt Google-Maps mit Echtzeit-Daten manches fest verbaute Navigationssystem schlecht aussehen. Dieses angespannte Verhältnis wird künftig mit intelligenten Assistenten nicht einfacher. Waren es bislang hauptsächlich Apple (Apple CarPlay) und Google (Android Auto), sowie kleinere Anbieter wie Nuance mit Dragon Drive, drängt nun auch Amazon mit einer Autoversion von Alexa in den Markt. Das Unternehmen kann mit ersten Integrationen bei Audi, Seat, Ford, Lexus, Lincoln und Toyota punkten. In Auto-Use-Cases schiebt sich der Assistent vor alle anderen Funktionen und greift dabei tief in das Fahrzeug ein. Im Zweifel kann man mit dem Auto-Assistent nicht nur Einkäufe erledigen und Musik, Podcasts sowie anderen „Radio-typischen“ Content abrufen, sondern auch die Kernfunktionen des Fahrzeugs selbst (Navigation, Licht, Scheibenwischer usw.) steuern. Entsprechend intensiv ist das Ringen um dieses Feld – und die Zahl der Kombinationen, in denen die Autohersteller in unterschiedlichen Märkten mit verschiedenen Plattformen kooperieren, ist verwirrend und faszinierend zugleich. Während BMW in den USA mit Apple und Google geht, wird im chinesischen Markt natürlich AliGenie integriert. Google bietet mit Android Automotive ein komplettes Betriebssystem für moderne Fahrzeuge. Neben der Installation weiterer Apps im Auto kann der Fahrer via App tief in die Fahrzeugfunktionen von außen einsehen bzw. eingreifen. Der Polestar 2 von Geely/Volvo wird im Jahr 2020 mit Android Automotive auf den Markt kommen. Bleibt die Frage: Wem gehören die generierten Daten? Wer darf daraus neues Geschäft generieren?

Der Anwendungsfall Auto ist so attraktiv, dass sich nicht nur die großen darum bemühen.

Auch kleinere Unternehmen und Start-ups sind hoch aktiv. Mit Byton kommt 2020 eine neue Elektroauto-Marke auf den Markt. Das Unternehmen hat seine Wurzeln in China, verfolgt jedoch einen internationalen Ansatz. Der große Bildschirm (125 cm breit), der das gesamte Armaturenbrett im SUV M-Byte ausmacht, wird über Gesten, Touchpads und auch Sprache gesteuert. Als intelligenten Assistenten nutzt Byton in China die Lösung von Baidu und in Europa Amazons Alexa.

Aus Berlin stammt das Unternehmen German Autolabs, das mit „Chris“ ein kleines Gerät (siehe Abb. 6.2), vergleichbar mit einem kleinen Navigationsgerät, wie TomTom einst war, für Endkunden auf den Markt bringt, in dem der gleichnamige Assistent Chris einige Sprachfunktionen ermöglicht. So erhält der Nutzer Zugriff auf die Navigationsfunktion seines Handys, kann sich WhatsApp-Nachrichten vorlesen lassen. Chris agiert dabei kontextabhängig und stellt beispielsweise während des Passierens von Kreuzungen keine Anrufe durch. German Autolabs bietet die in Chris zugrundeliegende Technologie auch im B2B-Kontext für Flottenfahrzeuge an.

Der Überblick über die Marktteilnehmer darf die vielen und vollkommen verschiedenen, meist eher kleinen Anbieter nicht außer Acht lassen, die zum Teil hoch spezialisierte Lösungen in Kombination aus Hard- und Software anbieten. So zum Beispiel [Elliq.com](#) aus Kap. 3, die einen intelligenten Assistenten für ältere Menschen anbieten. Die Idee einer Betreuung von Senioren durch Maschinen, mag zunächst etwas seltsam und befremdlich erscheinen. Der Smartspeaker wirkt sehr abstrakt. Sein kopfähnliches Element besticht durch ein von innen beleuchtetes „Auge“. Bedenkt man jedoch die massiven Herausforderungen, die eine alternde Gesellschaft in westlichen Industrienationen mit



Abb. 6.2 Chris ist für den Einsatz in Autos konzipiert. (Foto: Hersteller)

sich bringt, können intelligente Assistenten hilfreich sein. Die Zahl der zu pflegenden älteren Menschen wächst schneller als die Zahl der Pflegekräfte. Ein niedrigschwelliges Sprachinterface mag nicht die abschließende Lösung des Dilemmas in der Altenpflege bedeuten, aber es kann einen wertvollen Beitrag leisten. Ältere Menschen kommunizieren, bleiben mit dem Internet verbunden und auf diese Weise vernetzt. Die dadurch entlasteten Altenpfleger können sich Arbeiten widmen, die von Robotern auf lange Sicht nicht vernünftig abgedeckt werden können. Vor diesem Hintergrund ist der Slogan von Elliq durchaus passend: „Elliq, der Sidekick für glücklicheres Altern“. Elliq ist ein KI-gesteuerter Sozialbegleiter, der ältere Erwachsene wach, vernetzt und engagiert halten soll. Elliq ist das erste proaktive KI-Gerät, das Interaktionen mit ihren Benutzern initiiert und mit Hilfe von maschinellem Lernen personalisiert und an seine Benutzer angepasst wird.

Einen ähnlichen Ansatz, wenngleich auch auf eine vollkommen andere Nutzergruppe zielend, verfolgt der bereits im Kap. 2 vorgestellte Sprachlehrer [Lily.com](#). Mit endloser Geduld arbeitet der Sprachtrainer ebenfalls in Form eines Smartspeakers mit Antromorphismus-Elementen, einem stilisierten Kopf und Rumpf, der an einen menschlichen Körper erinnert.

Zum Schluss der Betrachtung der Marktteilnehmer und Marktakteure soll auf eine Gruppe hingewiesen werden, die möglicherweise für viele noch nicht im Bewusstsein ist: die Regulierungsbehörden. Auf das Spannungsfeld zwischen dem Wunsch, dass CUIs und intelligente Assistenten Aufgaben autonom erfassen und ausführen mit dem dafür notwendigen Vertrauen, das Nutzer den Anbietern dieser Dienste und Geräte entgegenbringen, da die Funktionen ja eher intransparent und für den Einzelnen schwer nachvollziehbar sind, haben wir bereits hingewiesen. Je stärker intelligente Assistenten Eingang finden und sich nützlich machen, desto wahrscheinlicher ist es, dass dies nicht immer optimal gelingt, dass sich dabei Interessenslagen zwischen Anbieter und Nutzer nicht immer sauber abgrenzen lassen und mithin auch Missbrauch und Schadensfälle auftreten werden. Dies ist ein Problem, das weit über das Thema dieses Buches hinausreicht und natürlich auf alle Arten von Einfluss durch künstliche Intelligenz zutrifft. Es ist aber wichtig zu erkennen, dass eben nicht nur die Anbieter von Software und Hardware, nicht nur die Dienstleister und Agenturen diesen Markt bestimmen, sondern mit hoher Wahrscheinlichkeit auch Regulierungsbehörden in die Abläufe eingreifen werden.

Literatur

1. Samsung: The bixby Marketplace is now officially open in the U.S. and Korea, (01.07.2019), <https://news.samsung.com/global/the-bixby-marketplace-is-now-officially-open-in-the-us-and-korea>, zugegriffen 30.10.2019
2. Fahrenbach, Achim: Dieser Mann will Amazons Alexa zerstören – mit einem Sprachassistenten, der eure Privatsphäre schützt, (04.12.2018), <https://www.gq-magazin.de/auto-technik/article/dieser-mann-will-amazons-alexa-zerstoeren-mit-einem-sprachassistenten-der-eure-privatsphaere-schuetzt-rand-hindi>, zugegriffen 23.10.2019
3. Adobe, Funktionen von XD., <https://www.adobe.com/de/products/xd/details.html>, zugegriffen 30.10.2019



Look East

7

Zusammenfassung

In Europa schauen wir fast ausschließlich auf die Angebote von Google und Amazon, Apple und Microsoft. Doch gerade in asiatischen Ländern hat die Entwicklung von CUIs bereits kräftig Fahrt aufgenommen. Die Gründe: Demografie und technologisches Leap-Frogging. Etliche technische Entwicklungsschritte, die wir im „alten“ Europa erst gehen mussten, werden in Asien einfach übersprungen. Zudem existieren in den Ländern vorwiegend Silbensprachen, die auf Tastaturen schwer darstellbar sind. Einfach mit den Geräten zu sprechen, löst dieses Dilemma und erschließt Bevölkerungsschichten, die zwar ein Smartphone besitzen, aber Schwächen im Lesen und Schreiben haben. Hinzu kommt, dass die Akzeptanz neuer Technologien in Asien generell höher ausfällt als in Europa.

Wenn es um digitale Innovationen geht, wenden wir für gewöhnlich den Blick nach Westen. In den USA, insbesondere an der Westküste im Silicon Valley und Seattle werden Produkte entwickelt, die wir täglich benutzen. Im Fall von CUIs und intelligenten Assistenten zählen zwar auch die amerikanischen Unternehmen zu den Innovationstreibern, allerdings ist die Entwicklung in Asien vielleicht noch ein wenig dynamischer und die Services sind weiter entwickelt. Wer CUIs und intelligente Assistenten verstehen will, muss nach China, Indien und auf andere asiatische Märkte schauen.

7.1 Demografie und Sprachen

Dass CUIs und intelligente Assistenten in Asien so bedeutend sind, hat gleich mehrere Ursachen: die Demografie, die Sprachen und den Verlauf der digitalen Entwicklung. Auf diese Aspekte soll in diesem Kapitel eingegangen werden:

Das Durchschnittsalter asiatischer Bevölkerungen ist – mit Ausnahme von Korea und Japan – deutlich jünger, als dies in Europa der Fall ist. Während die Europäer durchschnittlich 43 Jahre alt sind, beträgt das Durchschnittsalter in Indien gerade einmal 28 Jahre und in Indonesien 32 Jahre. Selbst China ist mit einem Durchschnittsalter von 37 Jahren erheblich jünger als Europa (zum Vergleich: USA 38 Jahre, Deutschland, Japan 47 Jahre, Korea 41 Jahre). Jüngere Bevölkerungen gehen offener und neugieriger mit Innovationen um. Das kann man in verschiedenen Sektoren beobachten. Es gilt aber eben auch für die Adoption von digitalen Innovationen. Im Zusammenspiel mit dem Umstand, dass die technische Entwicklung der meisten dieser Länder erst in den 1980er- und 1990er-Jahren begonnen hat, kann man heute an vielen Stellen so genannte Leap-Frogging-Effekte feststellen. Bestimmte Entwicklungsschritte, die in den reiferen westlichen Volkswirtschaften durchschritten wurden, werden in Asien einfach übersprungen. Die Anzahl derjenigen, die in Indien das Internet durch einen Desktop-PC kennengelernt haben, ist verschwindend gering. Zum Zeitpunkt der Markteinführung von PCs stellten derartige Geräte selbst im professionellen Zusammenhang eher die Ausnahme dar, und im privaten Rahmen waren sie aus Kostengründen praktisch nicht vorhanden.

Für die meisten Bewohner Chinas, Indiens, Indonesiens und anderer asiatischer Märkte ist das Internet ganz selbstverständlich eine mit dem Smartphone verbundene Einrichtung. Wie man dies im Übrigen auch im Westen bei Kindern und Jugendlichen beobachten kann, die in der Zeit nach Einführung von Smartphones (nach 2007) geboren wurden.

Wie bereits dargelegt wurde, gibt es eine ganze Reihe von Gründen, weshalb das Smartphone als Plattform für natürlichsprachige Interfaces naheliegend ist. Einer dieser Gründe ist die Begrenztheit des Bildschirms, der gleichzeitig Ausgabe von Inhalten und Eingabe über die Tastatur leisten soll. Tastaturen müssen daher zwangsläufig kompakt sein. Da das heute gebräuchliche grafische Benutzerinterface von Menschen erfunden und weiterentwickelt wurde, die hauptsächlich mit dem in westlichen Sprachen gebräuchlichen lateinischen Alphabet umgehen, ist das Interface auch auf diese Notationsform optimiert.

Viele asiatische Sprachen nutzen aber Alphabete, die eigentlich Silbenschriften sind, oder ursprünglich Silbenschriften waren. Die Anzahl der Zeichen in diesen Notationsformen ist bedeutend höher. Es wird gesagt, dass die chinesische Schrift über einhunderttausend Zeichen kennt. Selbst, wenn man davon ausgeht, dass im umgangssprachlichen Gebrauch „nur“ dreitausend Zeichen gebräuchlich sind, handelt es sich dabei um ein Vielfaches der 26 Zeichen, die das lateinische Alphabet kennt – selbst, wenn man alle Sonderzeichen und Akzente berücksichtigen würde. Die meisten der über 20 indischen Sprachen werden ebenfalls in Alphabeten notiert, die aus Silbensprachen entstanden.

Tamil, eine Sprache, die immerhin von geschätzt 200 Millionen Menschen in Indien gesprochen wird, nutzt 247 Schriftzeichen. Es ist nachvollziehbar, dass die Notation derartiger Sprachen auf kleinen Bildschirmtastaturen unbefriedigend ist. Die Eingabe mittels gesprochener Sprache stellt natürlich eine große Erleichterung dar. In vielen asiatischen Märkten ist zudem die Analphabetenquote hoch oder die Anzahl derjenigen Nutzer, die zwar theoretisch Lesen und Schreiben können, dies aber nicht gerne tun. Sie bevorzugen das Sprechen und Hören. Demografie, Entwicklung des mobilen Internets und die Komplexität der asiatischen Sprachen sind wesentliche Treiber für die rasante Entwicklung, die natürlichsprachige Interfaces in China, Indien, Indonesien als auch in Vietnam, Thailand und anderen asiatischen Wachstumsmärkten vorantreiben.

7.2 Bocksprünge

Als Leap Frogging (Bockspringen) bezeichnet man das Auslassen einzelner Stufen im Laufe eines Entwicklungsprozesses. Asiatische Märkte haben das Desktop-PC-Zeitalter übersprungen und sind gleich auf mobile Geräte umgestiegen. In diesem Zusammenhang wurde in den meisten Fällen auch das gute alte Tasten-Mobiltelefon übersprungen und das damit gewohnte Versenden von SMS-Kurznachrichten. Die Funktion der SMS wurde in China beispielsweise sofort von Messenger-Diensten wie WeChat angeboten, die zwar ähnlich wie SMS durch das Eingeben und Empfangen kurzer Textnachrichten funktioniert, was die Adaption von Messenger-Diensten wie WhatsApp im Westen auch verlangsam hat, da viele Nutzer den Zusatznutzen von Messenger gegenüber SMS nicht sofort verstanden. In China und anderen asiatischen Märkten dagegen war Messaging sofort ein großes Thema und die vielfältigen Zusatzdienste, wie das Verschicken von Emojis, Bildern und weiteren Medieninhalten wurden unmittelbar Teil des normalen Kommunikationsverhaltens. Ebenfalls Teil des mobilen Lebens wurde Mobile Payment. Auch dies ein echter Leapfrog, denn der Siegeszug von Mobile Payment, das unmittelbar auch Teil von Messaging-Diensten wurde, konnte stattfinden, weil die Mehrzahl der neuen Mobile-Nutzer zuvor entweder gar keine Bankverbindung oder lediglich eine Bankverbindung mit schlechtem Service hatte. Mobile Payment stellte den Anschluss an die Neuzeit dar.

Wenn ein Wallet oder ein ähnlicher Dienst auf einem Smartphone installiert ist, kann Bezahlung und damit echte Transaktionen die Bedeutung von Messenger-Diensten erheblich steigern. WeChat ist damit zu einer der meistgenutzten digitalen Plattformen gewachsen. Alleine im ersten Quartal 2019 zählte es mehr als 1,1 Milliarden täglich aktive Nutzer. Dabei werden – anders als bei Facebook – bei einem großen Teil der Nutzungsvorgänge echte Transaktionen getätigten. So lassen sich in China und Indien über Conversational User Interfaces und mithilfe von intelligenten Assistenten vielerorts Bustickets, Lebensmittel und viele Dinge des täglichen Bedarfs auswählen und kaufen. Das ist ein Vorgang, an den sich westliche Nutzer mit Amazon Alexa gerade erst allmählich gewöhnen.

7.3 Boomende Annahme

Geht es um China, geht es um Wachstum. Das ist auch bei CUIs nicht anders. Chinesische Unternehmen setzen massiv auf Sprache und bringen Smartspeaker und andere Geräte, die mithilfe von Sprache steuerbar sind, auf den Markt.

Alibaba verkaufte seine TMall Genie Speaker, ein Gerät vergleichbar von Amazon Alexa, im Rahmen des jährlichen chinesischen Shopping Festival „Singles' Day“ für umgerechnet 15 US-Dollar. Die anderen Unternehmen wie Baidu, Xiaomi und Huawei folgten sofort. In der Folge hat sich ein Trend verstärkt, der bereits im Jahr zuvor deutlich sichtbar wurde: Im Jahr 2017 war der Smartspeaker-Markt in China weltweit noch vollkommen unbedeutend und Amazon hatte auch daher 80 Prozent Anteil am Weltmarkt. Bereits Anfang 2018 war der Anteil von Amazon deutlich geschrumpft. Einerseits weil Google seinen Assistenten in den Markt gebracht hatte, vor allem aber auch, weil Alibaba, Xiaomi und andere eine sehr dynamische Markteinführung im Riesenmarkt China schafften (siehe Abb. 7.1). China hat sich aus dem Stand zu einem Schwerpunktmarkt für Smart Speaker entwickelt. Ein Markt, in dem derzeit weder Google noch Amazon vertreten sind.

Chinesische Unternehmen entwickeln erfolgreich vollständige Ökosysteme, wie beispielsweise Xiaomi zeigt. Im Westen ist das Unternehmen vor allem als Anbieter von Smartphones und Tablets bekannt, die stylisches Design mit guter Performance zu einem aggressiven Preis anbieten. Immer wieder wurde das Unternehmen mit dem Bildzeichen „MI“ als iPhone-Schreck bezeichnet. Im Westen arbeitet das Unternehmen daran, diesem Ruf gerecht zu werden. In China sieht die Sache ganz anders aus. Dort genießen die Smartphones einen sehr guten Ruf. Xiaomi betreibt eigene Shops, in denen eine Vielzahl von Tech-Gadgets verkauft werden, die allesamt natürlich über die Smartphones und den darin integrierten intelligenten Assistenten XIAO steuerbar sind. Derart ausgestattet gibt es Raumüberwachungskameras, Steuerung von Klima und Temperatur, Remote-Türöffner

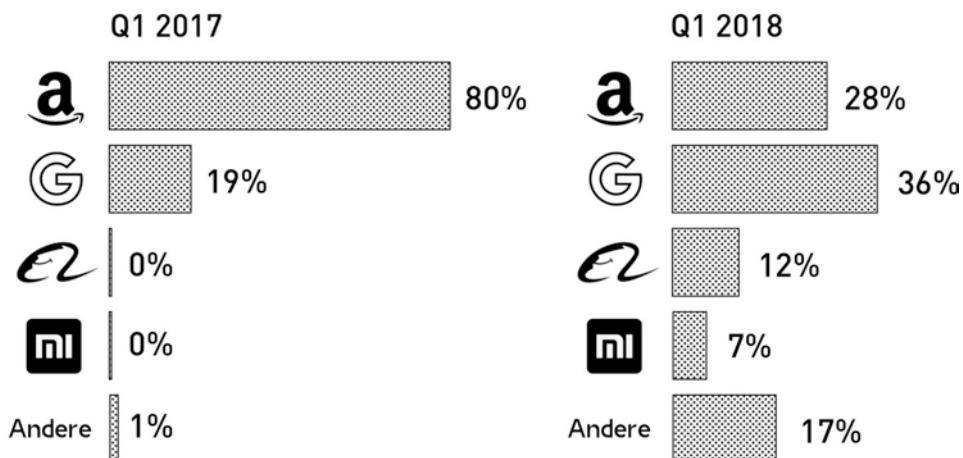


Abb. 7.1 Verschiebung der Marktanteile von CUIs in China

und viele Home-Automation-Geräte, die zusammen ein Ökosystem bilden. Xiaomi hat zwar kein eigenes Betriebssystem, wie dies bei Google und Apple der Fall ist, kombiniert aber geschickt Android Open Source und eigene Komponenten. Damit bietet Xiaomi für Nutzer und Entwickler ein attraktives Netzwerk an. Für Entwickler ist interessant, dass Xiaomi eine Art Crowdfunding-Plattform, wie Indigogo, betreibt und in besonders erfolgversprechende Gadget-Entwicklungen einen sehr aktiven eigenen Beteiligungsarm hat, der wie ein aggressiver Hardware-Software-VC agiert.

Wie sehr diese Entwicklung bei der aufsteigenden Mittelschicht Chinas ankommt, ist aus europäischer Perspektive häufig irritierend. Ein neues Lebensgefühl macht sich breit, bei dem der Einzug der künstlichen Intelligenz ins tägliche Leben und die Nutzung von digitalen Diensten selbstverständlich ist. In der U-Bahn von Shenzhen in China kann man Menschen beobachten, die zu ihrem Smartphone sprechen und dabei offenbar mittels Videoüberwachung von Raum zu Raum wechseln, schließlich bei Haustier, Kind oder Großmutter landen, um dann vollkommen selbstverständlich in ein Gespräch mit diesen Anwesenden im Haus überzugehen. Europäische Sicherheitsbedenken oder Hemmungen in Bezug auf Privatsphäre sind in diesem Teil der Welt und in dieser jungen Demografie nicht weit verbreitet. Im Gegenteil, es scheint ein Sicherheitsgewinn zu sein, die Kinder mit Bodycams auszustatten, die man jederzeit per Sprachnachricht aktivieren kann, biometrische Profile allerjenigen zu speichern, die durch die Wohnungstür treten oder am Ende des Tages mit einem Sprachassistenten Gebete an die Hausgeister zu übermitteln.

7.4 Indische Sprachvielfalt

Unter den am schnellsten wachsenden Volkswirtschaften der Welt und mit der zweitgrößten Bevölkerungszahl ist der indische Markt von großem Interesse für Internet- und Softwareunternehmen.

Bain & Companys umfangreicher Bericht „Unlocking Digital for Bharat (Bharat = das ländliche Indien)“ [1] schätzt, dass Indien heute 390 Millionen Internetnutzer hat, von denen jeder Fünfte ein Smartphone besitzt. Dabei handelt es sich weitgehend um eine junge, männliche, städtische Bevölkerungsgruppe. Ungefähr 900 Millionen Menschen sind noch keine Internetnutzer, nicht zuletzt weil sie mit dem grafischen Benutzerinterface nicht sozialisiert sind und teilweise auch funktionale Analphabeten sind. Conversational User Interfaces haben das Potenzial, den Online-Zugang zu einem großen Teil der indischen Bevölkerung zu erweitern und damit Zugang zu Informationen, Inhalten und Dienstleistungen aller Art zu gewähren.

7.5 Mehrdeutigkeit und Komplexität der Sprachen

Im Rahmen des Google Next-Billion-Plans wurde die Sprachsuche bereits in acht indischen Sprachen gestartet, damit Verbraucher ihre Stimme für Suchanfragen verwenden können.

Obwohl Englisch die Amtssprache Indiens ist, sprechen nur zehn Prozent der Inder Englisch. Neunzig Prozent sprechen Sprachen wie Hindi, Marathi, Gujarati, Bengali, Kannada, Telugu, Tamil, um nur einige der 29 Hauptsprachen zu nennen, die in Indien gesprochen werden. Eine Sprache zu verstehen bedeutet, Wörter, Phrasen, syntaktische Formen und Konzepte zu kennen und diese Konzepte auf sinnvolle Weise miteinander zu verknüpfen. Dies erfordert umfangreiche Kenntnisse über die Sprachen und die Fähigkeit, diese zu interpretieren. Viele der indischen Sprachen sind nur in geringem Maße in zugänglichen schriftlichen Dokumenten gefasst. Das gilt für die Dokumentation der grammatischen Regeln, aber noch mehr für das Vorhandensein umfangreicher digitaler Bibliotheken in diesen Sprachen, mithilfe derer die NLP-Algorithmen und Verfahrensweisen trainiert werden könnten. Wenn es notierte Dokumente gibt, sind die Notationen nicht besonders gut standardisiert, was es in Verbindung mit diesen Sprachfamilien immanenten nuancenreichen Bedeutungen sehr komplex macht, gut funktionierende Natural-Language-Bearbeitungsverfahren zu entwickeln.

7.6 Mangel an sprachlicher Grammatik, Literatur und dokumentierten Standards

Eine der größten Herausforderungen in Indien ist heute der Mangel an Ressourcen in Bezug auf Literatur und Grammatik, obwohl Millionen von Muttersprachlern diese Sprachen verwenden. Das Erstellen von NLP-Algorithmen ohne eine grundlegende lexikalische Ressource ist eine große Herausforderung. Es gibt regelbasierte Methoden, die sprachspezifisch sind, aber fehleranfällig. Das indische Ministerium für Elektronik und Informationstechnologie hat bei all diesen Bemühungen eine Initiative gestartet, um zumindest die 22 verfassungsmäßig anerkannten Sprachen im Unicode-Standard zu vertreten und zu dokumentieren.

All dies stellt nicht nur die indische Regierung vor große Herausforderungen, die gleichzeitig aber auch sehr gute Chancen der Inklusion von häufig noch immer ausgesetzten Bevölkerungsgruppen bietet, die mit den Computern, Smartphones und der englischen Sprache häufig überfordert sind. Chancen, die nicht nur in Indien, sondern auch in einer ganzen Reihe weiterer asiatischer Länder genutzt werden könnten. Überall forschen und entwickeln Computer- und Sprachwissenschaftler mit großen, teilweise transnational finanzierten Projekten Übersetzungssysteme ihrer jeweiligen Sprache in andere asiatische Sprachen – am häufigsten natürlich in Chinesisch. China ist nicht nur eine Sprachsupermacht, sondern vor allem die beherrschende asiatische Wirtschaftsmacht. Gewissermaßen im Windschatten von Übersetzungen entwickeln sich auch andere Felder der natürlichen Sprachverarbeitung. Hier ist wiederum Indien eines der Weltzentren der Kombination von Sprache und Machine Learning sowie *anderen* Verfahren. Während China eine besondere Rolle im Bereich automatischer Bilderkennung einnimmt, das zu einer der Schlüsseltechnologien für alle Formen autonomer Mobilität und Robotik gehört, nimmt Indien eine vergleichbare Stellung im Bereich des Natural

Language Processing ein. Eine ganze Reihe von indischen Top-Universitäten haben bereits vor längerer Zeit damit begonnen, sich mit den Problemen automatisierter Sprachverarbeitung zu beschäftigen. Möglicherweise auch dadurch motiviert, dass das Land eben eine große Sprachvielfalt hat, die jeder Inder ständig erleben kann und viele Inder dazu zwingt, wichtige Fragestellungen in einer Zweitsprache oder dem indischen Englisch zu organisieren. In der indischen NLP-Community werden regelmäßig wichtige Fragestellungen diskutiert und es wird an Lösungen gearbeitet, die auch für andere Regionen der Welt, die viele verschiedene, teilweise eher kleinere Sprachen pflegen – wie beispielsweise Europa – künftig von großer Bedeutung sein können. In der blühenden indischen Start-up-Szene von Hyderabad, Bangalore oder Neu Delhi gibt es eine große Zahl von Teams, die Lösungen unterschiedlichsten Zuschnitts anbieten. Vom ausgewachsenen intelligenten Kaufassistenten, wie ihn Niki.ai anbietet (siehe Kap. 2), bis zur kleinen Nischenlösung eines Gnani.ai, die Text-to-Speech-Module erstellt. Mithilfe dieser Module kann man verschiedene indische Sprachen in Text verwandeln, um daraus weitergehende Analysen bezüglich des Inhalts durchführen zu können.

Literatur

1. Sheth, Arpan; Bhattacharya, Joydeep: Unlocking the Digital Opportunity in India, (14.08.2018), <https://www.bain.com/insights/unlocking-the-digital-opportunity-in-india/>, zugegriffen 03.11.2019



Tech Stack: Wie die erfolgreiche Umsetzung im Unternehmenskontext gelingt – und warum das gesamte Unternehmen betroffen ist

Zusammenfassung

In diesem Kapitel liefern wir Umsetzungshilfe aus der Praxis. Harald Behnke, Director Consumer Experience (CX) Strategy EMEA bei Oracle, arbeitet mit Kunden, Partnern und Produktentwicklung an digitalen Innovationsthemen. Ein Schlüsselbereich zu Beginn der 2020er-Jahre ist die Transformation zu Conversational User Interfaces für alle Bereiche der Interaktion zwischen Kunden und Unternehmen. Der Praktiker beschreibt detailliert wie ein CUI im Unternehmen geplant, entwickelt, implementiert und zum Leben erweckt wird. Dabei geht es nicht nur um den Tech Stack mit Clients, Channels, Platforms und Backends. Der CX-Experte prophezeit, dass ein CUI auch das Denken und die Abläufe innerhalb des gesamten Unternehmens verändern wird.

„Alexa, frag mein Unternehmen \$IRGENDWAS“, beschreibt perfekt den Erkenntnisverlauf der meisten Nutzer und Anbieter von Sprach- und Chatbots: Mit den ersten erfolgreichen Konversationen stellt sich unwillkürlich die Frage, wann alle weiteren Dienste des Anbieters so bequem „erfragt“ werden können. Schnell wird klar: Chatbots sind nicht einfach ein weiterer Kanal, sondern das „dritte Internet“ nach Web und Mobile. Die komplexen GUIs (Graphical User Interfaces) werden zu CUIs (Conversational User Interfaces) – radikal einfach für den User, aber technisch komplex für die Anbieter. In diesem Sinne beschreibt dieses Kapitel, was Unternehmen und Institutionen benötigen, um in der Welt von intelligenten Assistenten und CUIs bestehen zu können.

Beitrag von Harald Behnke

Zunächst wird dazu die CUI Technologie-Landschaft mit dem Zusammenspiel der verschiedenen Schichten und Anbieter vorgestellt – von globalen Messaging-Diensten bis zu Plattformen im Unternehmen. Darauf aufbauend wird aus Sicht eines Unternehmens oder von Institutionen beschrieben, welche technologischen Herausforderungen die erfolgreiche Entwicklung eines massenmarktfähigen CUI in dieser Landschaft mit sich bringt. Schließlich beschreibe ich die Erfolgsfaktoren für eine erfolgreiche Projektumsetzung: Planung, Entwicklung, Distribution und Betrieb vom CUI-Pilotprojekt bis zur unternehmensweiten CUI-Transformation.

8.1 Technologielandschaft

Einen ersten Chatbot zu bauen, ist technisch recht einfach – einen nützlichen Chatbot dagegen sehr schwer. Denn die zu beherrschenden neuen Technologien sind komplex auf vier Schichten verteilt – mit jeweils wichtigen Teilaufgaben: Es gibt den Client, also die Geräte und Anwendungen über die der Endnutzer Gerätedaten ein- und ausgibt; den Channel, also den Messaging- und Assistenzdienst für Text und Sprachkommunikation; die Plattform auf der die Systeme zur Entwicklung und zum Betrieb von CUI, Bots und Assistenten bereitgestellt werden und schließlich das Backend, in dem funktionale Systeme Antworten liefern und Aufgaben übernehmen (siehe Abb. 8.1).

Schauen wir uns diese vier Schichten etwas genauer an, um später darauf aufbauend CUI-Konzepte zu entwickeln.

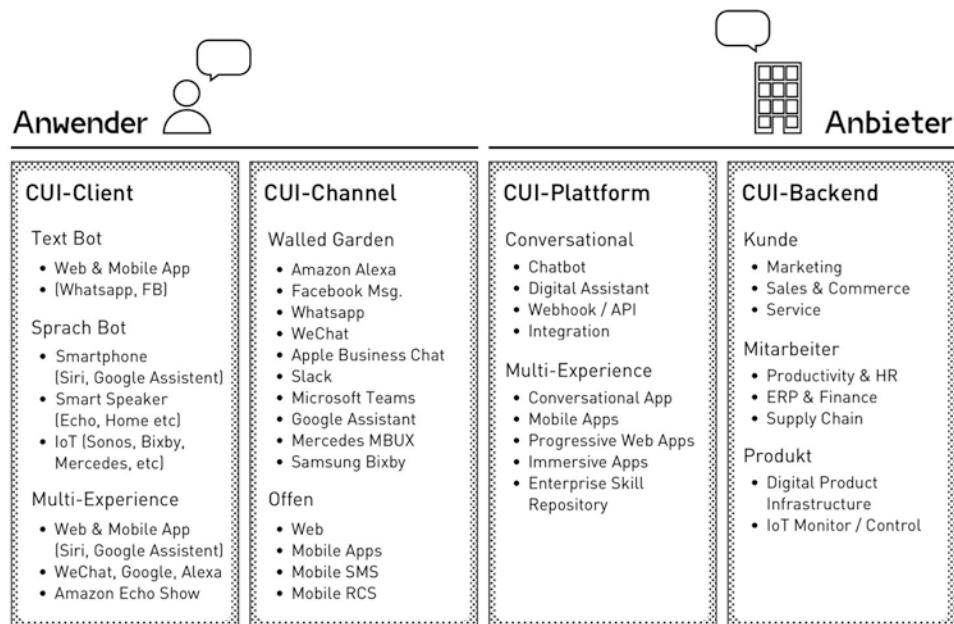


Abb. 8.1 CUI-Technologielandschaft

8.2 CUI-Clients und -Endgeräte

Der Erfolg von Chatbots basiert auf Milliarden Konsumenten mit Smartphones, die nun erhebliche Zeit in Messaging-Plattformen von Facebook Messenger, WhatsApp bis WeChat verbringen. Der Wettbewerb um dieses Massenpublikum führte zu vielen Technologieinnovationen von der Betriebssystem-Integration (Apple Siri) bis zu reinen Sprach-Endgeräten (Amazon Echo) – und der Öffnung der Ökosysteme für Chatbot-Entwickler von Drittunternehmen.

Anhand der für uns Nutzer „anfassbaren“ Endgeräte und Apps lassen sich die grundlegenden Arten von Conversational User Interfaces am besten beschreiben:

1. **Chatbot:** vom User initiiertes Einzweck-Gespräch („Taxi für 13 Uhr buchen“)
2. **Digitale/Virtuelle Assistenten:** vom System oder User initiierte Mehrzweck-Gespräche; in Betriebssystem oder Anwendungen kontextabhängig integriert („Zeit zum Aufbrechen für Deinen nächsten Termin“)
3. **Voice Assistants:** Verlagerung der Konversation auf natürliche Sprache
4. **Multi-Experience Assistants:** Kombination von Text und Sprache mit multimedialen und interaktiven grafischen Benutzerschnittstellen ermöglicht gerade auf Smartphones schnellste Ergebnisse
5. **IoT-Assistants:** Sprache wird zunehmend zur Steuerung von digitalen Produkten genutzt – z. B. Mercedes MBUX Voice Assistant, Samsung Bixby, oder Google, Alexa und Siri zur Kontrolle von Unterhaltungs- und Haustechnik.

8.3 CUI-Channel (Messaging- & Digital-Assistant-Dienste)

Ein gutes Medien- und Technikverständnis der unterschiedlichen und in Kap. 6 bereits näher beschriebenen Kommunikationskanäle ist entscheidend für das Erreichen der gewünschten Zielgruppe und Gesprächsziele. Zwar ist die Kernfunktion überall ähnlich: die Übertragung der Konversationen, Adressierbarkeit der Nutzer sowie die Übersetzung von Sprache in Text und umgekehrt. Doch gibt es erhebliche Unterschiede in Bezug auf Qualität, Quantität, Offenheit sowie Sicherheit der verschiedenen Chatbot-fähigen Kanäle:

- **Quantität:** Wie viele Personen meiner Zielgruppe erreiche ich über welchen Kanal und damit verbundenen Endgeräten? Und erreiche ich die Zielgruppe mit gleichem Funktionsumfang? So hat der Marktführer WhatsApp zwar 60 Millionen Nutzer in Deutschland, ermöglicht aber weder interaktive Card-Module wie Google Chat (20 Millionen Nutzer), sehr offenen Zugang wie Facebook Messenger (31 Millionen Nutzer) noch gesprochene Interaktion wie das weit weniger verbreitete Amazon Alexa.
- **Qualität:** Welchen CUI-Funktionsumfang erlauben die in meiner Zielgruppe genutzten Kanäle? So ermöglicht die eigene Unternehmenswebseite oder mobile App höchste CUI-Sicherheit, hier sind z. B. Bank-Transaktionen möglich, die bequem über Alexa Sprachsteuerung derzeit noch nicht funktionieren. Für interne Chatbots dagegen ist der

Funktionsumfang von Slack und Microsoft Teams recht ähnlich. Ein weiteres Qualitätsmerkmal ist die Spracherkennung und Ausgabe bei Siri, Alexa, Google oder Samsung. Sie unterscheidet sich teils erheblich, weil die Entwicklung von Natural Language Processing (NLP) und Natural Language Understanding (NLU) sich noch in einem frühen Stadium befinden [1].

- **Offenheit:** Chatbots und CUI-Anwendungen können nicht überall gleich distribuiert werden. Im offenen Internet, also beispielsweise über die eigene Website, über mobile Apps und SMS können Bots weitgehend ohne Einschränkungen bereitgestellt werden – allerdings fehlen die Komfortfunktionen der Walled-Garden-Kanäle. Für Entwickler weitgehend offen sind hier z. B. Facebook Messenger, Amazon (Alexa Skills) und Google Assistant (Actions) mit vielen Zusatzfunktion wie Sprach-/Textumwandlung. Erst seit kurzem hat sich Marktführer WhatsApp über die Business API für Chatbots geöffnet, hat jedoch ähnlich wie Apple Business Chat einen restiktiveren Genehmigungsprozess. IoT-Produkt-Kanäle wie Mercedes MBUX sind für externe Chatbot-Entwickler in der Regel noch verschlossen.
- **Sicherheit:** Auch hier unterscheiden sich die verschiedenen Chatbot-fähigen Kanäle für Nutzer und Anbieter: Sind Identitäts- und Altersprüfung möglich? Wie erfolgt sicherer Zugang zu Backend-Systemen im Unternehmen? Entsprechen Privatsphäre und Datenresistenz den gewünschten Standards (DSGVO (Datenschutzgrundverordnung), CCPA (California Privacy Act) oder darüber hinausgehende Anforderungen). Bei einigen globalen Messaging-Plattformen werden Behörden und Finanzwirtschaft Risiken sehen. Aber auch viele Unternehmen scheuen sich, wichtige Daten, die durch gewissermaßen intimen Austausch mit Nutzern natürlich entstehen, leichten Herzens mit den großen Anbietern zu teilen. Andersrum lassen diese US-Anbieter zum Schutz Ihrer Endnutzer Chatbots aus regulierten Industrien (Pharma und Medizin, Tabak und Spirituosen) nicht zu – obwohl der Betrieb über Web oder SMS erlaubt wäre. Eine wichtige Rolle als besonders sicherer Kanal für Chat Messaging und intelligente Assistenten könnte mittelfristig der SMS-Nachfolger Standard RCS (GSMA Rich Communication Services [2]) übernehmen. Ähnlich wie bei Telefonie und SMS handelt es sich um eine dezentrale, regulierte Infrastruktur auf Basis offener Standards. Daher könnten die Probleme von zentralen, nicht-regulierten Messaging Plattformen überwunden werden. Durch die native Unterstützung im Google Android Betriebssystem – Anwender brauchen so keine Apps mehr für Chat und intelligente Assistenten – sowie verstärkter Netzbetreiterunterstützung in den USA und Europa könnte RCS nach langjähriger Anlaufphase ein relevanter Kanal werden. Die zwei ausstehenden Meilensteine zum Durchbruch von RCS sind die fehlende Unterstützung von Apple iOS und macOS – die GSMA hofft, dass Apple iMessage [3] zukünftig RCS statt wie heute SMS als Fallback-Kanal nutzt – sowie die in der Praxis bewiesene Interoperabilität zwischen verschiedenen Netzbetreibern.

8.4 CUI-Plattformen

Die CUI-Plattformen spielen eine zukünftige Schlüsselrolle im Unternehmen, da ein Großteil der Innovationen und Investitionen für die CUI-Transformation – neben der Integration oder Modernisierung der verbundenen Backend-Systeme – hier erfolgen. Die Kernaufgaben der CUI-Plattformen sind

- a) die Verbindung zu allen Kanälen,
- b) die Steuerung des Gesprächsfluss,
- c) das Routing und die Integration mit allen Backend-Systemen und
- d) die Bereitstellung von Konfigurations- und Entwicklungs-Frameworks, Bot-Vorlagen und Betriebswerkzeugen.

Aufgrund der Komplexität der menschlichen Sprache verwenden CUI-Plattformen Machine Learning, NLP und NLU – weshalb Conversational UI oft auch Conversational AI genannt wird. Die Grundlagen der Sprachverarbeitung wurden bereits in Kap. 4 erläutert, und die wesentliche Aufgabe einer CUI-Plattform ist, diese Komplexität für den praktischen Einsatz im Unternehmensalltag zu vereinfachen. Die Grundfunktionen einer CUI-Plattform sind jedoch recht einfach zu beschreiben und folgen der Gesprächs-Metapher:

- **Intelligenter Assistent** (Chatbot, Agent): Die für den Nutzer erlebbare End-Anwendung, welche aus einem oder mehreren Bots sowie integrierten Kanälen und Backend-Systemen bestehen kann.
- **Bots/Apps** (Amazon Skills, Google Actions, Bixby Capsules, Agenten Software: Mini-Programme oder Anwendungen mit der Gesprächslogik für einen eng abgegrenzten Themenbereich, welches beschreibt, wie die eingehenden Äußerungen analysiert, verarbeitet, an andere Systeme weitergeleitet und abschließend beantwortet werden. Abhängig vom Konzept können Bots vollständig isoliert (Endnutzer-Bot-Repository) oder miteinander verbunden entwickelt werden (Enterprise-Bot-Repository).
- **Utterance**: Eingehende „rohe“ Äußerung vom Nutzer als Text, die gegebenenfalls zuvor vom sendenden Kanal oder der Plattform von Sprache in Text umgewandelt wurde. („Hallo, ich brauche einen Wagen für um fünf“)
- **Intent**: Erkennung der grundlegenden Absicht in der Nutzeräußerung auf Basis vorher eingegebener, trainierter oder ML-ermittelter Phrasen („Taxibestellung“)
- **Entity, Context & Expectations**: Erkennung von Parametern („17:00 Uhr CET“), Kontext („Taxifahrt“, „User 45883“) und Erwartung („Bestellung“) in der Intent-Außerung zur variablen Aufbereitung für weitere Systeme
- **Flow Routing**: Steuerung des Gesprächsflusses von der Weiterleitung an Backend-Systeme bis zur Ausgabe an den Nutzer.

- **Response:** Senden einer Antwort über den richtigen Kanal, der den Text gegebenenfalls wieder in Sprache umwandelt oder ergänzende interaktive oder multimediale Element darstellt. („Ja gerne, der Wagen 472 ist in 5 Minuten da“)
- **Bot Repository:** Mehrere Bots (Apps, Skills) können serverseitig ähnlich wie Apps auf dem Smartphone gesammelt und verwaltet werden, um Interaktionen auf mehrere Themenbereiche auszuweiten – sowohl vom Endnutzer im Kanal (Alexa Skills, Google Actions) als auch vom Unternehmen in der CUI-Plattform (z. B. Oracle Skill Repository mit Backend-Integrationen). Master-Bot-Systeme können mehrere Bots (Skills, Apps) zu Mehrzweckanwendungen kombinieren.

Das Marktforschungsunternehmen Gartner zählt aktuell 26 Anbieter für CUI-Plattformen: 16 Anbieter für spezialisierte „Conversational Platforms“ [4] und 13 für ganzheitliche „Multi-Experience Platforms“ [4]. Einige Anbieter bieten kanalgebundene Plattformen an (z. B. Google, Amazon), andere fokussieren sich auf Teilaufgaben oder erweiterten ein Backend-Silo-System (z. B. Avaya, Genesys, Servicenow). Nur IBM, Microsoft und Oracle decken beide Kategorien ab.

Conversational Platforms fokussieren sich auf die Entwicklung und Betrieb von Chatbots und Virtuelle Assistenten. Anbieter in dieser Kategorie sind z. B. Google, Rasa, Microsoft oder Oracle. **Multi-Experience Platforms** erweitern das Spektrum auf mobile Apps, progressive Web Apps, Conversational Apps und immersive Apps. IBM, Kony, Microsoft, Oracle, SAP oder Salesforce bieten entsprechende Lösungen.

8.5 CUI-Backend

In vielen Fällen liefert die Antwort auf eine Nutzerfrage nicht die CUI-Plattform, sondern wie bei diesen Beispielen ein externes Backend-System: „Was kostet der Flug nach Rom?“, „Wie ist mein Bestellstatus?“, „Schließe die Fenster meines Autos!“. Es gibt im Wesentlichen zwei Ansätze für die Integration von Backend-Systemen mit der CUI-Plattform:

1. **API/SDK Integration:** Diese verbreitete Integrationsmethode verbindet CUI-Plattformen und Backend über eine bei allen Systemen verfügbare, nicht vordefinierte Schnittstelle. Beispiele für diese Methode bieten nahezu alle CUI-Plattformen. Dabei werden Datenstrukturen des Backends über Programmierschnittstellen mit der CUI-Plattform verbunden, so dass die Gesprächs- und Geschäftslogik in der CUI-Plattform mit Daten des Backend-Systems arbeiten kann. Sämtliche Logik wird dabei von Grund auf neu in der CUI-Plattform entwickelt.

Vorteil ist die große Flexibilität und Offenheit. Nachteil sind die hohen Kosten und Zeitaufwände für den Umbau und Integration eines in der Regel mit grafischem Benutzer-Interface konzipierten Altsystems. Besonders geeignet ist die Methode bei einfachen Anwendungsfällen oder wenn das Backend-System wie bei Start-ups ebenfalls neu in der Entwicklung ist.

2. **Backend-Bot/Skill-Integration:** Bei dieser Methode sind CUI-Plattform und Backend-System durch Software-Module (Backend-Bots, Skills, Apps, Agenten) bereits beidseitig für Conversational Use Cases vorbereitet und integriert. Das Backend-System bringt dabei über die reine Datenanlieferung hinaus auch Gesprächsfunktionalität mit. Vorteile sind die geringeren Integrationskosten, höhere Sprachverständlichkeit und Themen-Bandbreite durch erleichtertes Multi-Purpose Skill Routing (Beispiel: Soll ein erkanntes Intent wie „Hilfe“ vom vorintegrierten Commerce, Marketing oder Service Backend beantwortet werden?), Best-Practice-Wiederverwendung (komplexe Geschäftslogik im Backend bekommt CUI-Funktionen, muss so nicht doppelt in der CUI-Plattform entwickelt werden) und schnellere Bereitstellung. Nachteil sind höhere Anschaffungskosten. Besonders geeignet ist diese Methode bei etablierten Unternehmen mit hohen Investitionen in bestehende Backend-Systeme. Ein Beispiel für diese neuere Methode ist der Oracle Digital Assistant mit Backend-Skills für unterschiedliche SaaS-Anbieter sowie Oracle ERP, HR und Customer Experience Cloud. Die Vorteile zeigen sich beispielsweise beim Chatbot des Logistikunternehmens Hermes [5], das sich aufgrund der Vorintegration von CUI-Plattform und Service-Backend und der damit verbundenen Sicherheits- und Geschwindigkeitsvorteile für Oracle entschieden hatte. Hermes-Kunden können so ihre Paketsendungen per Assistent in der mobilen oder Web App verfolgen und bearbeiten. Nach Einführung werden bereits 40 Prozent aller Kundenkontakte vom digitalen Assistenten beantwortet.

8.6 Technische Herausforderungen

Der Einblick in die CUI-Technologie-Landschaft zeigt zunächst zwei grundsätzliche Herausforderungen bei der Umsetzung. Erstens kommen in fast allen Bereichen junge Technologien wie UX Design, ML, NLP, NLU mit entsprechenden Lernkurven, Potenzial und Risiken zum Einsatz (siehe Abb. 8.2).

Zweitens müssen Investitionsstrategie und Technologiearchitektur – ähnlich wie ehemals bei frühen Webportalen und mobilen Apps – trotz zunächst unklarer ROIs berücksichtigen, dass CUIs langfristig das gesamte Unternehmen verändern werden. Konkret sind für die CUI-Konzeption und -Entwicklung aber vor allem fünf technische Herausforderungen wichtig (Abb. 8.2): Reichweite, Verständlichkeit, Intelligenz, Themen-Bandbreite und Relevanz.

8.6.1 Reichweite

Auf Basis der oben skizzierten Endgeräte- und Kanalrestriktionen sollte zunächst eine detaillierte technische Reichweitenanalyse für das CUI-Projekt definiert werden. Welcher gemeinsame Nenner an Chatbot-Funktionen über mehrere Kanäle und Geräte hat die größte technische Reichweite in der Zielgruppe? Benötigt das Projekt Sprache- oder

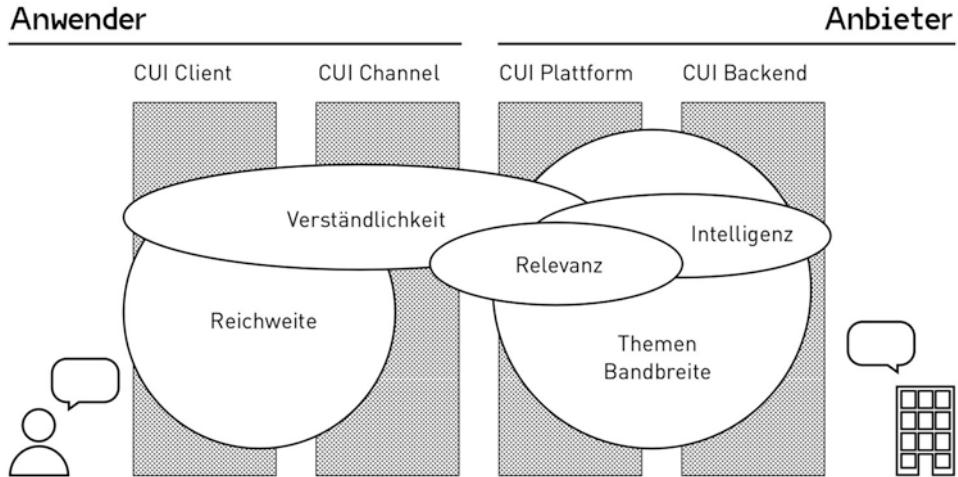


Abb. 8.2 Technische Herausforderungen für die Anbieter

Text-Konversation, sind visuelle Interaktionen oder sicherheitskritische Transaktionen notwendig? Diese Vorgaben zur Reichweite sind entscheidende Parameter für Projekt-ROI sowie CUI-Plattform-Auswahl.

8.6.2 Verständlichkeit

Schlechte Verständlichkeit eines Sprach- oder Chatbots wird von Menschen sofort erkannt und führt schnell zum Abbruch der Nutzung. Hohe Verständlichkeit eines CUI ist jedoch komplex und hängt von vielen Komponenten in der Prozesskette ab: der Qualität der Sprach/Text-Umwandlung in der Plattform oder im Kanal (Siri vs. Alexa), der NLP/NLU-Spracherkennung in der CUI-Plattform, dem Umfang der Skill-Programmierung (z. B. sprachliche Sonderfälle und spontane Benutzer-Korrekturen) sowie der Integrationsqualität mit dem Backend. Antworten können dementsprechend generisch vordefinierter Text sein, Elemente aus einer Knowledge-Datenbank mit unterschiedlichen Texten, ein synthetischer Text auf Basis von Nutzerdaten, eine Antwort mit Daten aus diversen Backend-Systemen – oder geplante Rückfragen zur Erlangung weiterer Daten. Aufgrund dieser nicht vorab kontrollierbaren Komplexität gilt es im Betrieb umfangreiche KPIs zu messen: Wie viele Konversationen wurden abgebrochen, welche Gesprächsflüsse und Intent-Muster sind weniger und mehr erfolgreich? Diese Messungen sind Basis für ein agiles Vorgehensmodell mit lernenden Systemen, die Verständigungsprobleme im Betrieb in der CUI-Plattform nachsteuern. Selbstverständlich entstehen alle digitalen Systeme in einem iterativen Prozess. Dies gilt aber für CUIs in noch viel höherem Maße, da die Nutzungspfade viel verschiedener sein können, als das bei vordefinierten GUIs naturgemäß der Fall ist.

8.6.3 Intelligenz

Die Nützlichkeit einer automatisierten Konversation entsteht vor allem aus dem vollständigen Domain-Verständnis und der funktionalen Tiefe von komplexen Backend-Systemen. Ob Flugbuchungen, Autovermietungen, Fernsteuerung von digitalen Services und Produkten, Automation komplexer Unternehmensprozesse und Daten: Umfassende, relevante, kompetente und schnelle Antworten auf alle Fragen im Kontext werden als intelligent *wahrgenommen*. Umgekehrt: fehlt diese „Intelligenz“ aus Sicht des Endnutzers, scheitern Chatbots als triviale Spielerei. Insbesondere bei CUI-Transformationsprojekten sind daher die Backend-Integration oder -Modernisierung die größten Investitionsbereiche.

Strukturellen Herausforderungen bei der Backend-Integration:

- **Domainspezifische Software & Daten** (z. B. ERP, CRM, Flugbuchung): Hier sind in der Regel semantische Metadaten sowie ein klarer Interaktionsfluss vorhanden, der jedoch für grafische Benutzerschnittstellen optimiert ist. Typische Probleme bei der GUI/CUI-Übersetzung sind reduzierte Spracheingabemöglichkeiten oder die Ausgabe von langen Listen. Lösungsansätze sind die unter CUI-Plattform beschriebenen Backend-Integrationsmethoden sowie der Ausbau von CUI-freundlichen Ein- und Ausgabeschnittstellen von GUI-Standardsoftware.
- **Domainunspezifische Software & Daten** (z. B. Content Management, Dateisysteme): Hier fehlen in der Regel semantische Metadaten und Interaktionsfluss, so dass die Entwicklung von CUIs keinen inhaltlichen Ansatzpunkt hat. Lösungsansätze der Softwareanbieter sind besser strukturiertes Knowledge-Management sowie intelligente Ratgebersysteme, die unstrukturierte Workflows und Daten in regel- oder KI-basierte Prozesse und Policies überführen und auch für CUI-Plattformen bereitstellen [6].

8.6.4 Themen-Bandbreite

Wie eingangs erwähnt, lösen multi-purpose Digital Assistant CUIs die single-purpose Chatbots der ersten Generation aus gutem Grund ab. Nur mit ausreichender Themen-Bandbreite ist Nutzerakzeptanz zu erreichen: spontane Kontextwechsel sind Basis *menschlicher* Konversationen – und jede alternativ bereitstehende Webseite deckt ja ebenfalls die gesamte Bandbreite eines Unternehmens ab. Für die Umsetzung ganzheitlicher Digitalassistenten muss eine entsprechende „CUI-Portal“-Architektur umgesetzt werden. Dabei empfängt ein Master Bot alle ein- und ausgehenden Äußerungen und leitet sie je nach Intent an die zuständigen speziellen Bots (Skills, Apps, Agenten). Sobald ein Skill eingeschaltete Intents nicht mehr versteht, kann über den Master-Bot ein Kontextwechsel erreicht werden. Über diesen Portalansatz können auch mehrere Backends des Unternehmens parallel angesprochen werden.

8.6.5 Relevanz

Was ist nun die *Kunst* bei der Umsetzung eines nützlichen digitalen Assistenten? Relevanz. Und die entsteht im geschickten Zusammenspiel aller vorher genannten technisch-konzeptionellen Faktoren sowohl im Kleinen als im Großen.

Mikro-Ebene: Im Kleinen ist eine richtige Balance aus Intelligenz, Empathie, Kontext und Geschwindigkeit wichtig für ein relevantes Gespräch. Eine Flugbuchung im Browser kann 30 Flüge auflisten – aber Alexa sollte vielleicht nur die günstigsten drei vorlesen. Und statt die KI alle Eventualitäten errechnen zu lassen, kann das System einfach öfter rückfragen: „Möchten Sie heute buchen oder recherchieren sie erst? Soll ich Ihnen Details per E-Mail senden? Fragen Sie bezüglich unseres Gesprächs von letzter Woche?“. Im Verlauf der Projektlaufzeit wird eine immer intelligentere Integration zwischen neuen Kanal-Komfortfunktionen, CUI-Plattform-AI und Backend-Skills diese Balance kontinuierlich verbessern. Eine zunehmende Bedeutung in Unternehmen spielt aus diesem Grund die zentrale Speicherung der Verhaltensdaten – das Nutzereinverständnis vorausgesetzt. Nur der gesprächslogikspezifische Teil dieser Daten sollte in der CUI-Plattform gespeichert werden. Darauf hinaus sollte die Datenspeicherung zentral und kanalübergreifend erfolgen, damit der Kunde per CUI, Web oder Mobile GUI oder vor Ort am POI im Kontext bedient werden kann. Da herkömmliche CRM-Systeme nicht für die Speicherung großer Mengen von Verhaltensdaten konzipiert wurden, entsteht für diese Anwendungsfälle die neue Software-Kategorie der sogenannten CDP (Customer Data Platform) und CIP (Customer Intelligence Platform) [7].

Makro-Ebene: Im Großen wird die Marke, Identität und Persönlichkeit eines digitalen Assistenten wichtig – und Technologie unterstützt dabei. „Alexa, frag XY nach einem London-Flug“ – ist Unternehmen XY auf Amazon Alexa überhaupt technisch präsent? Und warum sollte der Nutzer den XY-Chatbot überhaupt direkt ansprechen? Wie bereits in Kap. 3 dargestellt liegt ein Teil der Antwort in gutem Branding – der Chatbot-Name wird ebenso wichtig wie die .com-Webadresse. CUI-Plattform sowie UX-Konzeption können des Weiteren mit positiven Gesprächserlebnissen helfen. Sprechen zum Beispiel alle Backend-Subsysteme eines Konzerns mit gleicher, sympathischer Tonalität, werden Ihre Kunden geduzt oder gesiezt, sind die Begrüßungen und Phrasen über alle Abteilungen gleich formell, cool, flapsig oder frech? Ist ihre Gesprächsfluss-Programmierung in der Lage, mich zum Lächeln zu bringen? Technisch können hier zentralisierte Templates und Gesprächsmuster in der CUI-Plattform bei der Umsetzung helfen. In Kap. 9 werden weitere allgemeine Aspekte wie die explizite und implizite Adressierung und Auffindbarkeit von digitalen Assistenten weiter vertieft.

8.7 Projektumsetzung

Erfolgreiche digitale Assistenten zeigen, wie die genannten Herausforderungen durch entsprechende Planung, Umsetzung, Distribution und Betrieb bewältigt werden können. Allgemein sollte ein agiles Vorgehensmodell (junge Technologien erfordern Flexibilität) mit

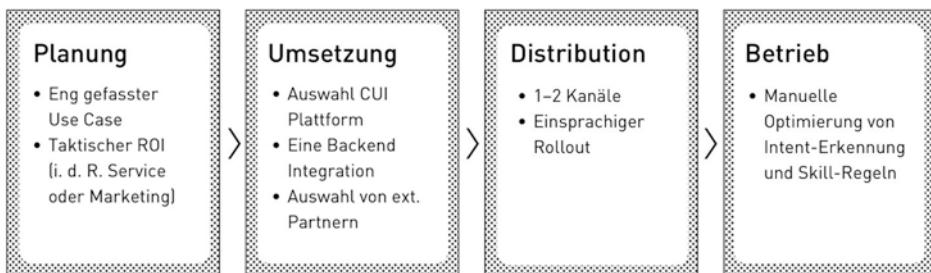
einer strategischen Plattformwahl (CUI betrifft langfristig alle Bereiche) kombiniert werden. Die weiteren Erfolgsfaktoren unterscheiden sich nach Projektart – von Pilot bis hin zu umfangreicher Transformation (siehe Abb. 8.3)

In Teilen wiederholt sich bei CUI-Projekten dabei die Geschichte des früheren Internets. So vermeiden Sie die Fehler von Web-Portal oder Mobile App beim neuen „Sprach-Portal“:

- Konzeption und Gestaltung von jungen Medien sind ein Engpass – sichern Sie sich früh Talente in neuen Berufsbildern wie Conversational User Interface Designer.

CUI Pilot

Ziel: Es soll gelernt werden, ob Kunden oder Mitarbeiter Interaktionen über CUI angenommen werden



CUI Transformation

Ziel: Aufgrund von Wettbewerbsdruck werden alle Angebote eines Unternehmens als CUI bereit gestellt, um Kunden nicht zu verlieren und Kosten zu senken

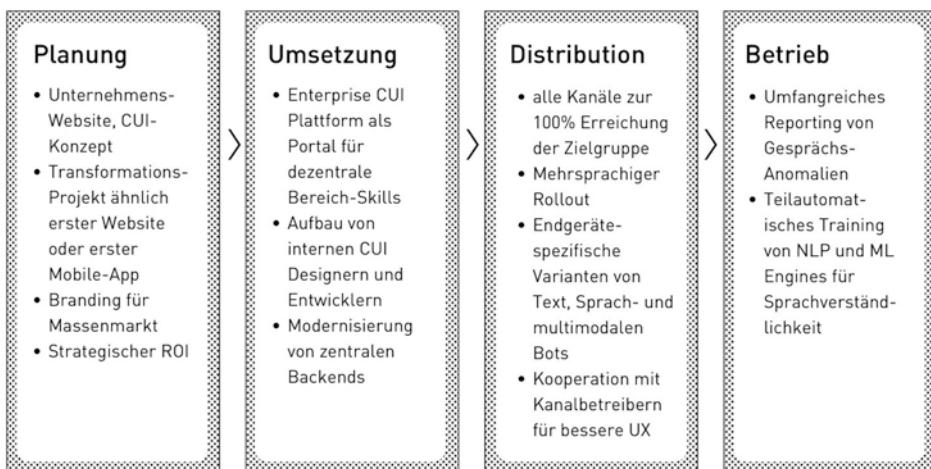


Abb. 8.3 CUI – Erfolgsfaktoren für die Umsetzung

- Die initiale „Tool-Gap“ für das entstehende Medium führt zu experimentellen, spielerischen Projekten – aber erst industrielle Plattformen und Werkzeuge (von CMS- zu CUI-Plattformen) erlauben komplexe Großprojekte rentabel zu realisieren. In der frühen Lernphase ist die Entwicklung von Einzelprojekten der Normalfall. Bei unternehmensweiten Projekten sind jedoch die Best-Practice-Wiederverwendung von Gesprächsmustern und Templates mit Hilfe von CUI-Plattformen aus wirtschaftlichen Gründen unumgänglich.
- Wie früher merken verschiedene Unternehmensbereiche erst zeitversetzt, dass sie „auch mit auf das neue (Web/Mobile App/CUI) Portal möchten“ – oder starten zum Nachteil des Nutzers parallele Silo-Angebote. Zeit zum Aufräumen: Wie viele Apps hat Ihr Unternehmen im Apple App Store? Nach wie vielen Skills auf Amazons Alexa werden Ihre Kunden fragen?
- Die globalen Kanalanbieter befinden sich in Standardschlachten – ähnlich zu Zeiten der „Browser-Wars“ können Nutzer heute noch nicht frei zwischen „Hey Google, Hey Siri, Alexa, Bixby“ wechseln. Um absehbare Kanalveränderungen nicht als Umbaukosten zu tragen, hilft wie damals genaues Beobachten – so hat Amazon gerade eine entsprechende Interoperabilitätsinitiative gestartet, die teure Mehrfachdistributionen von Bots vermeiden könnte [8].

8.8 CUI-Pilot

Zur erfolgreichen Umsetzung sollte zunächst ein eng gefasster Use Case geplant werden, der jedoch nützlich genug für den öffentlichen Einsatz ist, um negative Reaktionen über einen „Schrottbot/Crapbot“ zu vermeiden. Und damit muss wie oben beschrieben in der Regel zumindest ein Backend-System rudimentär integriert werden. Viele Unternehmen erreichen einen schnellen ROI durch Kostensenkung im Servicebereich oder Erschließung neuer Zielgruppen als Marketingziel.

Da der Pilot im Erfolgsfall ähnlich wie die erste Homepage eines Unternehmens weiterlebt, sollten Branding und Marketing für Namen und Persönlichkeit des Bots bereits weitgehend final umgesetzt werden. Die Entwicklung des Bots kann sehr gut an einen erfahrenen Partner ausgelagert werden, der Aufbau interner CUI-Experten ist noch nicht entscheidend. Die CUI-Plattform für die Umsetzung sollte jedoch unter eigener Kontrolle aufgesetzt werden und Potenzial haben, Unternehmensstandard zu werden: nicht nur Partner und Mitarbeiter lernen das System, auch die NLU- und ML-Systeme lernen im laufenden Betrieb und kumulieren ihre Erfahrungen.

8.9 CUI-Transformation

Die Planung einer CUI-Transformation sollte sich an allen Angeboten der Webseite orientieren. Welche Abteilungen müssen in die Planung einbezogen werden? Welche Backend-Systeme sind beteiligt? Sind die Backends voneinander unabhängig, oder

muss der Master-Bot tiefere Verknüpfungen zwischen den verschiedenen Backend-Skills organisieren? Sind einige Backend-Systeme strategisch so wichtig für den CUI-Erfolg, dass eine Modernisierung oder ein Austausch einer simplen Integration vorzuziehen ist?

Die Umsetzung erfordert dauerhafte Verbesserungen und Betrieb. Es sollte daher ein vollständiges internes Team [9] aus Designern, Entwicklern und Admins gebildet werden, das gemeinsam mit spezialisierten Dienstleistern an die Umsetzung geht. Bei allem Fokus auf Technologie kann nicht genug betont werden, wie entscheidend menschliche Kreativität und völlig neue Fertigkeiten für den Projekterfolg sind:

- **CUI-Designer:** Die Konzeption von Gesprächsschnittstellen ist Pionierarbeit und ähnelt dem Schreiben von Drehbüchern. Die meisten UI-Regeln und Best Practices werden mit dem Entstehen neuer Messaging-Plattformen erst entwickelt. Moderne CUI-Plattformen ermöglichen Content Authoring und Publishing für Nicht-Entwickler. Kernaufgabe ist, Aufgaben für den Nutzer so zu vereinfachen, dass alternative Kanäle wie mobile App oder Web nicht mehr vermisst werden.
- **CUI-Entwickler:** Die meisten CUI-Plattformen kombinieren proprietäre Frameworks mit klassischen Webentwicklungs-Script-Sprachen wie JavaScript. Die Kanal- und Backend-Integration sind Hauptaufgabe und kombinieren junge, testintensive Technologien. Beim Übergang von Entwicklung zu Betrieb ist abzuwagen, ob Verbesserungen der Spracherkennung durch manuelle Entwicklung oder teilautomatisiertes Machine-Learning-Training des Systems zu erreichen ist.
- **CUI-Admin:** Der erfolgreiche Betrieb hängt stark vom ständigen Training der Intent-Erkennungsmodelle ab, die manuell oder teilautomatisiert erfolgen können. Zudem hilft ein detailliertes operatives Reporting Designern und Entwicklern bei der Verbesserung von Spracherkennung und Gesprächsfluss.

Generell sollte der CUI-Betrieb wegen der hohen Investitionen in Backend-Integration sowie NLU und ML Lernprozesse von Anfang an zentralisiert werden. Neben der operativen Bereitstellung der CUI-Plattform gehört dazu auch das fortlaufende Reporting zur Erfolgskontrolle.

8.10 Wann ist ein CUI erfolgreich?

Wenn Unternehmen ein CUI im Einsatz haben, kommt auf die Verantwortlichen irgendwann die Frage zu: Wie bewerten sie den Erfolg eines Assistenten? Zählen allein die Installationen? Reicht schon, dass ein Nutzer einen Wunsch geäußert hat oder zählt nur ein abgeschlossener Arbeitsauftrag bzw. Kauf? Wie oft sollte ein Nutzer zurückkommen und erneut den CUI nutzen? Hier muss sich zunächst ein Gerüst an vergleichbaren KPIs entwickeln. Aber einige Kernthemen sind klar absehbar und sollten idealerweise mit den Analysewerkzeugen der CUI-Plattform ausgewertet werden können:

- Strategische Erfolgsmessung: Können angestrebte KPIs wie z. B. Call Deflection, Anzahl an Kundenkontakten, Verkäufe oder Neukundengewinnung erreicht werden? Kann der Marktanteil gegenüber Wettbewerbern durch Assistenten erhöht werden?
- Taktische Messungen für erfolgreicher Betrieb: Welche Kanäle und welche Unter-Bots sind am erfolgreichsten – von WhatsApp bis zur eigenen Webseite? Was sind die Top Intents bei abgebrochenen oder erfolgreichen Konversationen?
- Operative Messungen für fehlerfreien Betrieb: Welche Gespräche wurden abgeschlossen, welche abgebrochen? Welche Intent-Phrasen können durch Training verbessert werden? Welche ungeklärten Äußerungen führten zu keinem Intent? Gibt es systemische Fehler bei der Backend-Integration?

8.11 Fazit

Die neue Technologie der digitalen Assistenten zu meistern beginnt zunächst damit, sie nicht als „Chatbots“ zu unterschätzen – CUIs sind das dritte Internet. Wenn Sie sich mit einer unternehmensweiten CUI-Strategie spielerisch und agil in frühe Projekte begeben, die zudem auf einer zentralen Plattform nachhaltig wachsen können, werden Sie in dieser spannenden Ära nicht nur Erfolg haben – sondern sich nach einigen Jahren rückblickend fragen, wie man jemals anders mit Software interagieren konnte.

Literatur

1. Moren, Dan: Alexa vs. Google Assistant vs. Siri: Why Google Wins, (23.05.2019), <https://www.tomsguide.com/us/alex-vs-siri-vs-google,review-4772.html>, zugegriffen 02.11.2019
2. GSMA RCS - Rich Communication Services Standard, <https://www.gsma.com/futurenetworks/rcc/>
3. Apple: Information zu iMessage und SMS/MMS, <https://support.apple.com/de-de/HT207006>
4. Gartner Research: Market Guide for Conversational Platform, (30.07.2019), <https://www.gartner.com/doc/3953723>, zugegriffen 02.11.2019
5. Lindquist, Margaret: This Chatbot gets Thank You Notes, (13.8.2019), <https://www.forbes.com/sites/oracle/2019/08/13/this-chatbot-gets-thank-you-notes/>, zugegriffen 25.11.2019
6. Oracle: Process and Policy Automation, <https://www.oracle.com/applications/customer-experience/service/policy-automation.htm>, zugegriffen 03.11.2019
7. Bennett, Tess: CDP Market Heats Up With New Capabilities From Oracle CX, 19.09.2019, <https://which-50.com/cdp-market-heats-up-with-new-capabilities-from-oracle-cx/>, zugegriffen 04.11.2019
8. Amazon: Amazon and Leading Technology Companies Announce the Voice Interoperability Initiative, (24.09.2019), <https://press.aboutamazon.com/news-releases/news-release-details/amazon-and-leading-technology-companies-announce-voice?tag=theverge02-20>, zugegriffen 12.10.2019
9. Mozart, Franziska: Was macht eigentlich... ein Voice-User-Interface-Entwickler?, (01.06.2018), <https://www.lead-digital.de/was-macht-eigentlich-ein-voice-user-interface-entwickler/>, zugegriffen 25.11.2019



So verändern sich Ökosysteme

9

Zusammenfassung

Mit der Entwicklung weg vom Bildschirm hin zu (Laut-)Sprechern verändern sich auch die Ökosysteme im Internet sowie im Marketing. Die ständige Pendelbewegung zwischen Bundling (Bündelung) und Un-Bundling (Entbündelung) spielt eine wichtige Rolle. Informationen werden akustisch geliefert, ohne dass man mehr Webseiten aufrufen muss. Doch die Informationen stammen von Webseiten. Man kann also nicht auf diese Quellen verzichten. Gleichzeitig macht der Begriff „Voice only“ die Runde. Wozu noch Apps und Webseiten, wenn Sprache dominiert? Auch für das Marketing bringt die Entwicklung Herausforderungen: Wie platziert man eine Marke akustisch? Wie findet und entdeckt man neue Produkte, wenn man sie nicht sieht? Greifen Käufer weiterhin zum Markenprodukt, wenn sie per Sprache einkaufen oder akzeptieren sie auch No-Name-Alternativen? Das wird spannend, denn einer der größten CUI-Anbieter ist gleichzeitig eine Handelsplattform: Amazon.

Conversational User Interfaces begegnen uns an vielen Stellen im Alltag. Am bekanntesten dürften die Sprachassistenten in Smartphones oder Smartspeakern sein. Zunächst waren es allerdings Chatbots, die ihren Siegeszug vor allem über Instant Messenger wie ICQ oder Facebook Messenger angetreten haben. Sie nutzen natürlichsprachige Texteingabe und -ausgabe für ihre Dialoge und boten erste automatisierte Dienste an. Für eine breitere Öffentlichkeit dürfte 2011 Siri auf Apples iPhone der erste Kontakt mit einem sprachgesteuerten intelligenten Assistenten gewesen sein. Das Unternehmen war 2007 von Dag Kittlaus, Adam Cheyer und Tom Gruber im Rahmen eines Incubator-Programms des Unternehmens SRI International gegründet worden und seit Anfang 2010 als von Apple unabhängige App entwickelt worden. Wie Adam Cheyer in einem

Podcast-Interview [1] erklärte, mit dem Ziel den von Apples iPhone-Erfolg überraschten Wettbewerbern ein interessantes Produkt zu bieten. In dieser anfänglichen Version von Siri konnten Nutzer Anfragen in natürlicher Sprache stellen und erhielten auf intelligente Weise zusammengefasste Ergebnisse. Betraf die Frage beispielsweise eine Restaurantempfehlung konnte Siri die Ergebnisse vieler Restaurantempfehlungswebsites, wie beispielsweise Tripadvisor und Yelp, in einer intelligent sortierten Liste finden und auf Klick, sowohl eine Reservierung, als auch eine Taxiabholung (den Transportdienstleister UBER gab es zu diesem Zeitpunkt noch nicht) vom Standort buchen. 2011 kaufte Apple das Unternehmen auf und integrierte es mit dem Produktlaunch des iPhone 4S in das iOS-Betriebssystem, sodass jeder Nutzer eines iPhones oder iPads Siri nutzen konnte. Ohne gesonderte Installation, Schulung oder bestimmte Formulierungen konnte man Siri nach dem Wetter oder Auskünften fragen, oder einen Termin in den Kalender eintragen und den Timer beim Eierkochen per Stimmbefehl stellen. Der dabei ausgelöste öffentliche Hype übertraf alle Erwartungen seitens der Entwickler und auch Apples selbst. Der Börsenkurs von Apple entwickelte sich in den Wochen danach sehr gut, der Smartphoneanbieter stieg zum wertvollsten Unternehmen [2] der Welt auf. Siri war das erste massentaugliche Conversational User Interface, das die Möglichkeiten von automatisierter Sprachverarbeitung und intelligenter Assistenten einer breiten Öffentlichkeit vorführte. In der Folge gab es keine TV-Talkshow mehr, in der nicht Gäste Siri verrückte Fragen stellten („Siri, möchtest Du mich heiraten?“) und die Entwickler damit nicht nur vor ernste Kapazitätsprobleme stellte, sondern auch konzeptionelle Herausforderungen brachte. Die Machine-Learning-Funktionen konnten nicht schnell genug auf die unvorhergesehenen Fragen reagieren.

Diese große öffentliche Begeisterung war umso erstaunlicher, als die Apple-Version von Siri einen weit geringeren Funktionsumfang hatte, als die ursprünglich entwickelte Standalone-Version. Handelte es sich bei der App-Version um einen echten intelligenten Assistenten, der eigenständig Aufgaben erledigen konnte, war Siri als Teil des Betriebssystem eher eine Sprachsteuerung für die von Apple auf dem Smartphone vorinstallierten Apps, wie Kalender, Adressbuch usw. Ein Umstand, der wichtig ist, wenn man sich über die Entwicklung von Conversational User Interfaces und den damit verbundenen Ökosystemen Gedanken macht: Ist ein CUI lediglich Bestandteil eines Betriebssystems, oder ist das CUI selbst ein Betriebssystem?

Interessanterweise hatte Apple 1987 ziemlich genau 23 Jahre vor Launch von Siri als Teil des Betriebssystem bereits ein viel beachtetes Konzept als Video [3] vorgestellt: den Apple Knowledge Navigator. Der damalige Apple CEO Sculley erläuterte die Vision eines von Hardware unabhängigen intelligenten, sprachgesteuerten Assistenten, der multimodal sowohl in PCs, als auch in damals noch nicht auf dem Markt befindlichen Tablets oder auch vollkommen ohne Bildschirm nutzbar sein sollte. Er sollte in Echtzeit auf die damals noch unbekannte Cloud zugreifen können und durch künstliche Intelligenz Vorlieben und Nutzerkontakte in das Handeln einbeziehen. Vermutlich ist das visionäre Konzept des Apple Knowledge Navigators noch heute das Rollenmodell für viele Entwicklungen, die wir in diesem Buch beschreiben.

Ebenfalls viel Aufmerksamkeit erhielt 2015 die Einführung von Amazons Alexa. Der Sprachassistent erfuhr aufgrund des niedrigen Gerätepreises und Amazons Marktmacht schnell große Verbreitung. Alexa erkannte und setzte mehr Befehle um, als die Wettbewerber. Zudem öffnete Amazon die Plattform für Entwickler. Drittanbieter konnten Skills für den Assistenten anbieten. Die Idee ist nicht neu. Zum Erfolg des iPhones gehört die Entscheidung, einen App-Store mit Softwareprodukten von Dritten füllen zu lassen. Im Jahr 2016 zog Google mit seinem Assistant nach. Wie wir sehen, haben alle großen Technikabietere eigene Plattformen im Angebot oder sind dabei, solche zu entwickeln. Doch die Entwicklung von Conversational User Interfaces und intelligenten Assistenten steht noch ganz am Anfang.

Heute organisieren wir unsere Reisen mit dem Smartphone, überwachen große Maschinen aus der Ferne und nutzen Carsharing ohne feste Rückgabepunkte. Das Smartphone bzw. die darauf installierten intelligenten Apps sind aus unserem Leben nicht mehr weg zu denken. Doch begonnen hat es mit banalen Biertrinker-Apps (siehe Abb. 9.1). In den Anfangstagen hatte man die Lacher im Büro oder auf einer Party auf seiner Seite, wenn man so tat, als trinke man ein Bier aus dem iPhone. Zehn Jahre später ist das Ökosystem aus Hardware, Betriebssystemen, App und Drittanbieter-Software und den dahinter liegenden Backend-Strukturen international der vermutlich bedeutendste Ort digitaler Wertschöpfung. Im Feld von CUIs und intelligenten Assistenten stehen wir derzeit auf der Stufe der Biertrinker-Apps. Die Fähigkeiten von Smartspeakern und CUIs sind in vielen Fällen noch banal und es ist für viele kaum vorstellbar, wie daraus jemals etwas Relevantes entstehen sollte.

So wie sich die Welt seit den Biertrinker-Apps und Wasserwagen dramatisch weiterentwickelt hat und ein sogenanntes mobiles Ökosystem entstanden ist – ein Vorgang, der nur ungefähr zehn Jahre gedauert hat – so entsteht gerade ein Ökosystem rund um Conversational User Interfaces und intelligente Assistenten.

Abb. 9.1 Partyspaß in den App-Anfangstagen: Bier aus einem iPhone trinken (Foto: iBeer Megapack)



9.1 Die Ausdehnung des Universums

Die vielleicht wichtigste Dimension dieser Entwicklung im Feld der sprachgesteuerten Conversational User Interfaces ist bereits jetzt zu erkennen: Das relevante Ökosystem wird unendlich viel größer. Bislang ist das Internet nur aktiv nutzbar und zugreifbar in Form eines grafischen Benutzerinterfaces und mithin auch nur in den Situationen, in denen man Sichtkontakt herstellen kann und die Hände frei hat. Durch sprachgesteuerte Interfaces, kann man im Auto während der Fahrt einkaufen und beispielsweise Spülmittel nachkaufen, an das man sich gerade erinnert oder eine Kennzahl aus dem SAP-System abfragen, während man in den Zug nach Köln einsteigt.

Die Informationen und Dienstvielfalt des Internets werden über CUIs und intelligente Assistenten nahezu jederzeit und überall verfügbar. Die Anzahl und Andersartigkeit von Touchpoints wird für Anbieter, Unternehmen und Institutionen zu einer echten Herausforderung. Gleichzeitig aber selbstverständlich zu einer großen Chance, denn wenn es gelingt, das eigene Unternehmensangebot in dieser neuen Welt breit zu platzieren, wird der Kontakt zu Kunden und Nutzern natürlich intensiver. Anbieter erfahren deutlich mehr über ihre Kunden und können diese Erkenntnisse im besten Fall in Echtzeit in ihre Unternehmensentscheidungen und Angebote einbinden. Vermutlich wird es nur wenigen Unternehmen gelingen, an jedem denkbaren Touchpoint mit einem eigenständigen Angebot verfügbar zu sein. Große Unternehmen sind im Vorteil, aber selbst die größten Unternehmen sind in ihren Fähigkeiten begrenzt, weshalb die Unternehmen aus Silicon Valley, Seattle und China die Schwergewichte im sich entwickelnden Ökosystem sein werden. Durch die weitgehende Integration ihrer Dienste wie Messenger, Sprachverarbeitung, Knowledge Management, Payment und teilweise Hardware werden sie nicht leicht zu umgehen sein. Gleichzeitig ist bei der zu erwartenden Komplexität dieses sich entwickelnden Netzes aus Smartphones, Tablets, Wearables, Autos und anderen Verkehrsmitteln, Küchen- und weiteren Haushaltsgeräten, Haustechnik, TV und Game-Stationen nicht zu erwarten, dass dies alles von einem monolithischen Universal-Anbieter geleistet werden kann. Die Frage stellt sich, ob die derzeitig allmächtig erscheinenden Unternehmen tatsächlich den sich entwickelnden Gesamtmarkt beherrschen, wie dies im Smartphone-Ökosystem der Fall ist. Oder ob es sich um eine vorübergehende Übermacht handelt, wie auch AOL und CompuServe in einer bestimmten Phase des sich entwickelnden Internet-Ökosystems, eine vermeintlich uneinholtbare Dominanz entwickelt hatten. Die B2B-Sphäre in Büros und Produktionsumgebungen, in denen CUIs und intelligente Assistenten ebenfalls starke Veränderungen mit sich bringen, wird in dieser Betrachtung zunächst bewusst ausgeklammert.

9.2 Bundling und Un-Bundling

Es ist vielfach beschrieben, wie im Laufe der Digitalisierung ein ständiger Pendelschlag zwischen Bundling (Bündelung) und Un-Bundling (Entbündelung) von Diensten stattfindet. Gerne wird das Beispiel der Paketierung populärer Musik angeführt. Die Branche hat zunächst in Form von Singles einzelne Musikstücke vermarktet, dann eine Anzahl von

diesen Musikstücken auf einer Langspielplatte und CD an die Hörer brachte, um dann kurzzeitig zurück zu schwingen und einzelne MP3-Files vermarktet. Heute ist der Standard eher unbegrenztes Musikstreaming. Wie diese Pendelbewegung weitergehen wird, ist noch nicht abzusehen. Voraussichtlich gibt es Möglichkeiten, in der Auswahl und Vorstellung von Musikstücken und Interpreten wieder ein Bündel zu vermarkten, das dann auf Basis der in Spotify und Co verfügbaren Millionen von Songs Nutzern den Zugang erleichtert.

Eine durchaus vergleichbare Pendelbewegung gibt es auch in anderen Inhaltebereichen und Serviceleistungen. Die Aufgaben von Websites haben sich im Laufe der Jahre erheblich verändert und erweitert. Nicht nur wollen Websites ihren direkt zugreifenden Nutzern möglichst alle gewünschten Dienste und Inhalte möglichst übersichtlich liefern. Websites müssen zusätzlich auch die Suchmaschinen optimal bedienen, um bei den Websuchen gefunden zu werden. Dafür legen Websitebetreiber spezielle, für die Nutzer nicht sichtbare, Informationen in strukturierten Dateien auf den Servern bereit. Dies ist neben der optimalen Gestaltung der Website-Oberfläche und Inhalte ein wichtiger Teil der so genannten Suchmaschinenoptimierung (SEO). Wer Google als Suchmaschine nutzt, wird festgestellt haben, dass die Suchergebnisse immer häufiger nicht mehr nur eine reine Liste von Weblinks enthält, sondern in vielen Themengebieten einige Suchergebnisse in zusammengefasster Weise darlegt. Bei der Suche nach einem Nadelbaum wird der Wikipedia-Eintrag angerissen, bei Suche nach bestimmten Institutionen, die an einem bestimmten Ort lokalisiert sind, werden Öffnungszeiten und Adresse eingeblendet. Bei der Suche nach einer Flugverbindung von A nach B werden einige Flüge bereits in einer Weise angezeigt, dass der Nutzer diese durch einen Klick direkt bei einem Anbieter buchen kann. Anbieter, die in diesen Suchergebnissen derart gelistet sein wollen, müssen ihre Daten in besonderer Weise für Google aufbereiten, die der Suchmaschinenbetreiber mit ihnen abspricht. Dabei entnimmt Google in gewisser Weise den Inhalt des Anbieters und baut diesen als Modul in das eigene Interface ein. Ein neues Bundle ist geschaffen und die Website des eigentlichen Anbieters ist De-Bundled. Ein Vorgang, der all diejenigen aus dem Geschäft nimmt, beispielsweise Reiseportale, die bislang diese Art von Zusammenführung in Form von Marktplätzen oder Preisvergleichen eigenständig angeboten haben.

9.3 Von der App zur Micro-App

Der chinesische Dienst WeChat gilt in vielerlei Hinsicht als Rollenmodell für künftige Ökosysteme. Zunächst erscheint WeChat wie ein Messenger-Dienst – wie das westliche WhatsApp dazu gedacht, dass Nutzer untereinander Nachrichten austauschen, Gruppen bilden und darin kommunizieren. Wie bereits in Kap. 1 dargestellt, agieren bei WeChat auch Unternehmen, die eigene Profile betreiben, auf denen die Nutzer Dienste beziehen und Waren einkaufen können. Dazu genügen eine einfache Zustimmung, und sämtliche bei WeChat bereits hinterlegten Nutzerdaten und Zahlungsmittel können vom Drittanbieter genutzt werden. Anders als bei Apps der Smartphones, müssen diese Anwendungen vor

Nutzung nicht in einem App-Store heruntergeladen werden. Die entsprechenden WeChat-Funktionen von Drittanbietern stehen einfach im Messenger-Fenster bei Aufruf direkt zur Verfügung. Bedenkt man, wie ungern geübte Smartphonenuutzer heute noch Apps installieren – das Wort von App-Müdigkeit macht schon seit längerem die Runde – ist dies natürlich ein sehr nutzerfreundliches Angebot, das durch die integrative Plattformfunktion von WeChat ermöglicht wird.

In letzter Zeit haben sich neben den ausgewachsenen WeChat-Apps, die in der Regel den Chatverlauf im WeChat-Messenger unterbrechen und – wie westliche Apps auch – eine große Zahl von Funktionen über ein GUI anbieten, auch so genannte WeChat-Miniapps herausgebildet. In diesen stellen die Anbieter von WeChat-Apps einzelne Funktionen dieser Voll-Apps WeChat isoliert zur Verfügung. Auf diese Weise fragt der Nutzer nach einem Arzttermin in seiner Nähe und das Krankenhaus bietet sofort im Chatverlauf ein Terminvereinbarungsmodul an, welches man ohne weitere Registrierung und Probleme beim Verstehen des App-Interfaces nutzen kann. Rein – Raus. Ein De-Bundling der Voll-App findet statt und ein Re-Bundling im Chat-Fenster von WeChat. Der Vorteil für den Anbieter ist, dass der Nutzer schneller und mit niedrigeren Hürden sein Angebot wahrnimmt. Der Preis dafür ist, dass die Beziehung zum Nutzer weniger intensiv ist: WeChat Miniapps können keine Push-Nachrichten schicken, was traditionell ein wichtiges Werkzeug zur Nutzerbindung bei Apps ist. Mini-Apps haben auch andere Einschränkungen der aktiven Kommunikation in WeChat. So können Nutzer keine Posts in ihren Timelines posten. Miniapps müssen in einem bestimmten Format programmiert werden und Änderungen an Funktionen mit dem WeChat-Betreiber Tencent abgestimmt werden. Zudem sind sie weniger auf andere Plattformen portabel, als man dies von Apps kennt, die relativ einfach für IOS und Android aufbereitet werden können.

Auch im westlichen Smartphone-Ökosystem sind vergleichbare Tendenzen sichtbar. Während Standalone-Apps immer größer werden, sind sie häufig mit Funktionen überladen, in der Entwicklung immer teurer und am Ende nicht auf allen Smartphones sauber und performant. Zudem sind sie häufig für die Nutzer in den App-Stores nicht auffindbar und können den Endnutzer nur über massive Promotion dazu bringen, die App herunterzuladen. Das neue Modell heißt Progressive Apps, die den Unterschied zwischen Website, Web-App und Mobile-App verwischen. Oder Micro-Apps, die Google nach Suchen auf mobilen Geräten direkt aufruft und die in der Regel auch nur eine einzige Funktion anbieten (siehe Abb. 9.2).

Die Unübersichtlichkeit eines solchen Geflechts, in dem die Grenzen zwischen Betriebssystem-Plattform und Drittanbieter-Software (Apps, Skills, Actions) zunehmend verwischen, wird im Conversational User Interface und Intelligente-Assistenten-Ökosystem noch erheblich zunehmen. Die Microsoft-Suchmaschine Bing beispielsweise integriert in ihre Desktop- und Mobile-Suchergebnisse bereits seit Mitte 2017 zur Suchanfrage passende Chatbots, die man direkt in der Ergebnisseite in einem dazu aufspringenden Messenger-Fenster nutzen kann. Unternehmen und Institutionen werden versuchen müssen, gleichzeitig ihre eigenen Voll-Angebote im Desktop und App-Mobile-Umfeld zu

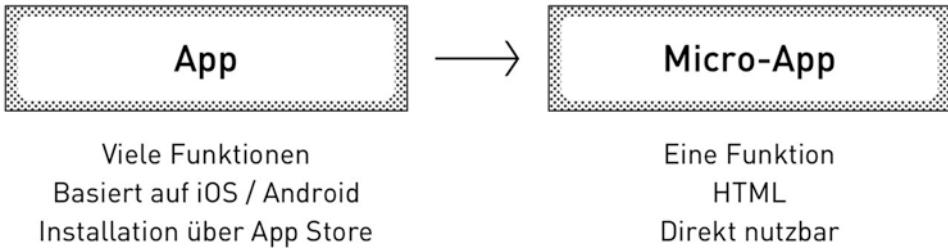


Abb. 9.2 Von der App zur Micro-App

betreiben und diese mit CUIs zur natürlichsprachigen Bedienung ausrüsten und ihre Inhalte und Dienste in viele kleine Teilstücke zerlegt und in Form von APIs und anderen Daten-Repositories paketiert den großen Plattformanbietern zur Integration in ihre Dienste zur Verfügung stellen.

Am Anfang des Kapitels haben wir bereits das außerordentlich visionäre Konzept von Apple Computer aus dem Jahr 1987 erwähnt. Grundüberlegung des Konzepts des Knowledge Navigators ist eindeutig, dass es sich um eine plattformunabhängige Agenten-software handelt, die sich auf einem offenen Marktplatz bewegt. Eine Struktur, die im vergleichsweise offenen Internet in Form von APIs irgendwie denkbar ist, aber nicht im für Assistenten und CUIs so wichtigen Mobile-Markt und dem Markt von Smartspeakern. Das damals visionäre Apple ist in diesen Tagen das genaue Gegenteil von Offenheit. Der Konzern und seine Produktwelt ist in seiner Hermetik sogar das Vorbild für vermeintliche Open-Source-Anbieter wie Google, die die interessantesten Funktionen und Möglichkeiten dem eigenen abgeschlossenen Teil des Betriebssystems vorbehalten und Lizenznehmer gar nicht oder erst viel später darauf zugreifen lassen.

Der bereits erwähnte Siri-Co-Gründer Adam Cheyer war als junger Mensch bereits Teil des Teams, das das Knowledge-Navigator-Konzept 1987 innerhalb weniger Wochen erdacht hatte. Nachdem er Siri an Apple verkauft hatte, gründete er das Unternehmen Viv Labs, das 2016 in einer spektakulären Live-Präsentation [4] in New York ein Software-Framework vorstellte, in dem Drittanbieter durch Bereitstellung entsprechender Schnittstellen einen offenen Marktplatz von Diensten, Inhalten und Services ermöglichten. Dieses demonstrative Gegenmodell zu den geschlossenen Gärten von Google und Apple konnte vor allem durch seine herausragende Fähigkeit, komplexe Bestellungen und Fragen auf verblüffende Weise zu beantworten. Eine Qualität, die offensichtlich mit der offenen Systemarchitektur erreicht wurde, die auf unterschiedliche Quellen autonom zugreifen konnte und die eigenen Handlungen und Algorithmen im Laufe des Gesprächs jeweils selbst on the fly programmierte. Viv Labs wurde Ende des Jahres 2016 von Samsung gekauft und offenbar seither in Samsungs Bixby-Assistenten integriert. Seit Mitte 2019 hat Samsung einen Marktplatz [5] vorgestellt, in dem offensichtlich wesentliche Ideen und Grundsätze der offenen Architektur von Viv aufgegriffen werden. Entwickler der Samsung-Bixby-Drittanwendungen (Capsule genannt), können ihre Anwendungen gleichzeitig enger in das Betriebssystem Bixby einbinden, Nutzerpräferenzen beziehen und gleichzeitig

mit anderen Drittanbietern teilen. Sie werden zugleich vom Assistenten Bixby viel stärker namentlich hervorgehoben, wenn Bixby mithilfe einer solchen Capsule eine Nutzerfrage lösen konnte. Der Siri- und Viv-Labs-Gründer Adam Cheyer arbeitet heute für Samsung.

9.4 Voice Search

Der Suchmaschinenbetreiber Google ist in einer wirklich einzigartigen Position. Nicht nur lagern auf seinen Servern ungeheure Schätze an Wissen über nahezu jedes Themengebiet der Welt, sondern über seine Suchmaschine erhält das Unternehmen täglich Milliarden Informationen darüber, was Menschen an bestimmten Orten zu bestimmten Zeiten interessiert und beschäftigt. Google erkennt, wie sich die Art der Suche und das dahinterliegende Bedürfnis der Menschen davon abhängig verändern, ob die Suche mit einer Tastatur auf einem Laptop, mit einer Bildschirmtastatur auf einem mobilen Gerät oder mithilfe der Stimme gestellt wird. Eine Sprachsuche ist regelmäßig weniger darauf bedacht, den Wunsch irgendwie in einem oder wenigen Stichworten für die Maschine zu abstrahieren, sondern ganz direkt: „Wie wechsele ich am schnellsten einen Autoreifen?“

Die Erwartung an die Antwort darauf ist ebenso direkt: Im Idealfall eine Schritt-für-Schritt-Anleitung, die der Nutzer auch mit schmutzigen Händen am Auto kniend Stück für Stück nachvollziehen kann.

Das ist keine einfache Aufgabe. Weder der Reifenwechsel selbst, noch die Identifikation des Bedürfnisses mittels Natural Language Processing, und schon gar nicht, die richtige Antwort in der richtigen Aufbereitung in der richtigen Sprache aufzuspüren.

All diejenigen, die sich mit der Erstellung von Inhalten, Meta-Informationen, Vertragungen und anderen Arten der Aufbereitung von Inhalten beschäftigen und dabei immer auch SEO als ihre Benchmark wahrnehmen, beginnt eine neue Zeit und die große Aufgabe: semantisches Denken. Dabei handelt es sich um eine radikale Form von nutzerzentriertem Arbeiten: Was möchte der Nutzer? Ausgangspunkt sind häufig die so genannten „W-Fragen“: Wer? Was? Warum? Wohin? Wann? Wie? Wo? (siehe Abb. 9.3) Die Formulierung der Inhalte sollte sich an diesen Fragen orientieren, um eine nützliche Basis für Sprachsuchen und die intelligenten Assistenten sein zu können.

9.5 Multiple Plattform

Da Google und andere Suchmaschinenbetreiber die Algorithmen ihrer Suchen und die Faktoren, die zu einem besseren oder schlechteren Ranking führen, nicht offenlegt, ranken sich um diese Faktoren seit jeher Spekulationen. Echte und vermeintliche SEO-Experten, SEO-Berater und SEO-Optimierer finden daher Aufmerksamkeit. Manche der Tipps sind simpel und gehören in den Bereich gesunder Menschenverstand (beispielsweise alle Informationen in maschinenlesbarer Form vorzuhalten, Telefonnummern,

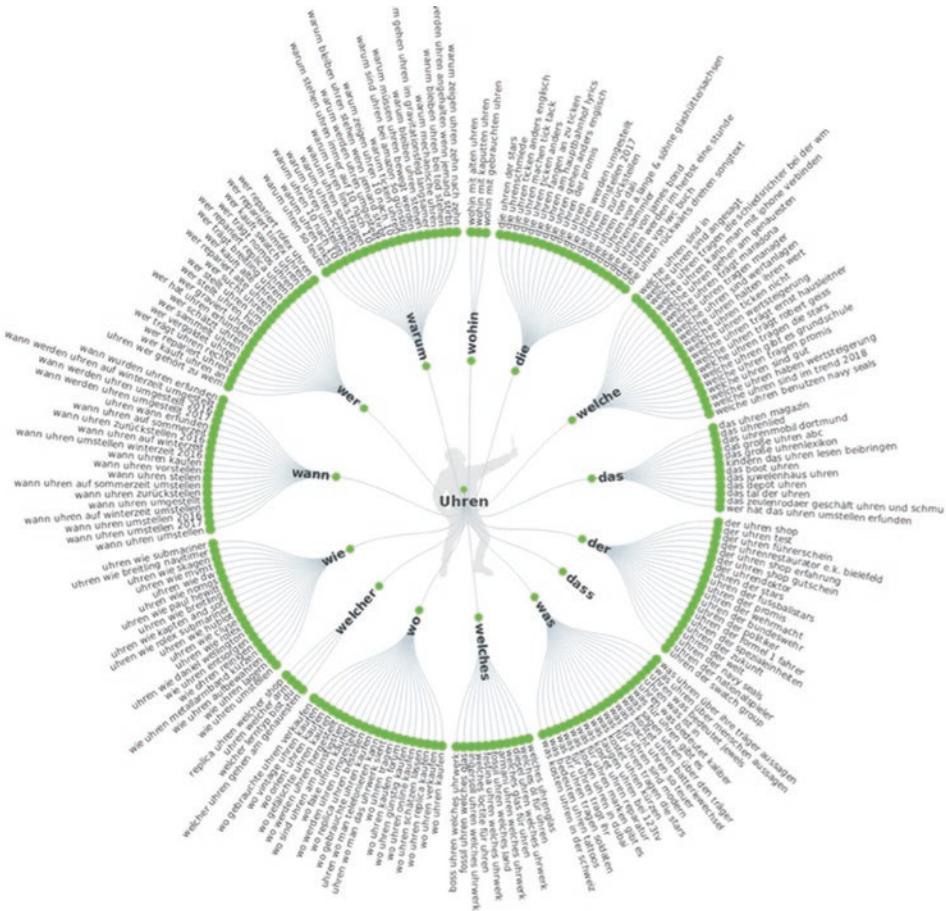


Abb. 9.3 Wort-Blume für alles rund um Uhren (Foto: AnswerThePublic.com)

Adressen, Öffnungszeiten für die Google-Suche lesbar zu machen usw.). Andere Dinge erfordern eher nutzerzentriertes Denken. Klar ist aber, dass die Art, wie Suchergebnisse präsentiert werden von Client und Endgerät abhängig sein müssen, auf denen diese präsentiert werden. Sind wir es an einen Desktop-PC gewohnt, unendliche Suchergebnisseiten zu durchforsten, schauen sich Nutzer normalerweise nur die erste Seite an. Viele glauben, dass auf Seite 2 der Google-Suchergebnisse bereits das so genannte Darknet beginnt. In einem Voice-only-Gerät, wie einem Google Home können wir in der Regel überhaupt nur ein Ergebnis, das uns vorgelesen wird, verarbeiten. Ein Verlesen von drei, vier oder gar zehn Ergebnissen würde den Usability-Rahmen vollständig sprengen. Ergebnisse auf Smartspeakern mit Display (z. B. Alexa Show) müssen wiederum anders gewichtet werden, als Ergebnisse, die auf einem Smartphone Display im Anschluss an eine Sprachsuche präsentiert werden (siehe Abb. 9.4). Anhand dieser Betrachtung kann

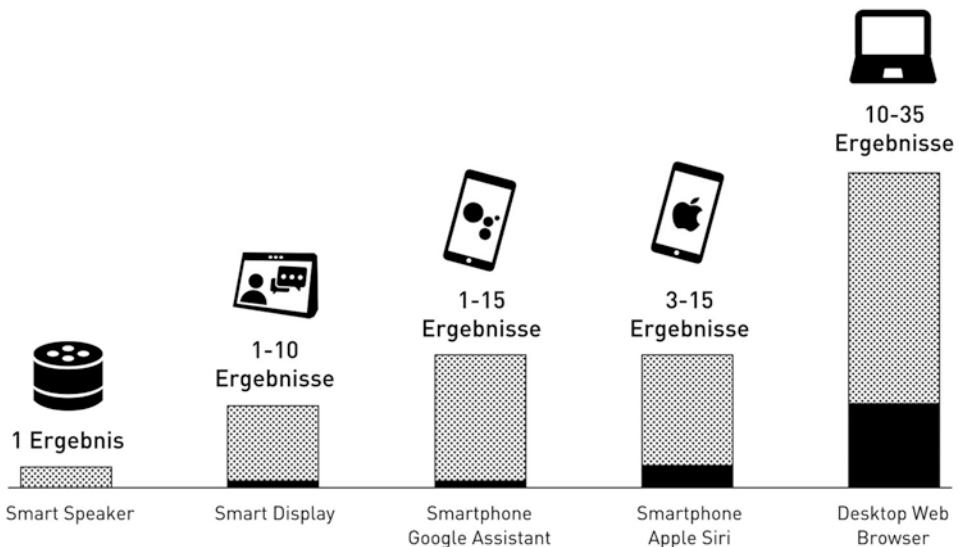


Abb. 9.4 Anzahl von Suchergebnissen in Gerät-Abhängigkeit (Foto: Voicebot.ai SEO Report, July 2019)

man gut verstehen, weshalb Conversational User Interfaces und intelligente Assistenten für Anbieter von Inhalten und Diensten strukturell engere Tore zum Nutzer darstellen, die auch noch von strikteren Torhütern kontrolliert werden.

9.6 Welche Bedeutung haben Websites künftig noch?

Vor dem Hintergrund der Überlappung von Voice Search mit Ausgabe auf einem Smartphone-Bildschirm und Voice Search mit Ausgabe auf einem Smartspeaker ohne Bildschirm oder einer Bedienung eines Conversational User Interface mittels Tastatureingabe in einem Messenger wird deutlich, dass Websites zwar noch ein responsives grafisches Benutzerinterface für Browser unterschiedlicher Größen und Anwendungsgebiete bereitstellen müssen, zunehmend aber Daten-Repositories werden, auf die Plattformbetreiber und Assistenten in unterschiedlicher Weise zugreifen können, um ihre jeweilige Content- und Service-Aggregation eigenständig vornehmen zu können. Bereits heute befindet sich die Suchmaschine Google mit Markttaggregatoren in unterschiedlichen Bereichen von Information bis Lead-Generierung im Wettbewerb und bei Suche nach einer Flugroute bietet Google als Suchmaschine, wie beschrieben, bereits ein vorbereitetes Feld an, in dem eine Reihe von möglichen Flugverbindungen auswählbar sind. An diesem Beispiel erkennt man, auf welche Weise die Disaggregation und Reaggregation von intelligenten Assistenten künftig funktionieren wird. Google gibt Anbietern von Inhalten und Dienstleistungen bereits heute Hinweise zur optimierten Strukturierung der jeweiligen

Angebote, damit die Suchmaschinen-Algorithmen diese gut einordnen können. Kürzlich kam eine weitere derartige Auszeichnung hinzu: speakable. Damit können Inhalte-Anbieter diejenigen Passagen ihrer Websites markieren, die sich besonders gut zum Vorlesen durch einen Sprachassistenten eignen. Auf diese Weise kann der Sprachassistent tief in die Ressourcen des Anbieters hineingreifen und in sein Sprachangebot integrieren.

Beim Aufbau neuer Online-Services denken manche Unternehmen schon darüber nach, inwiefern sie Voice first planen und handeln sollen. Der Begriff „Voice only“ macht beim Entwickeln neuer Apps die Runde, denn viele Unternehmen erkennen nicht nur, dass die Entwicklung einer Google Action oder Alexa Skill wichtig ist, sondern denken darüber nach, die eigenen Smartphone-Apps mit Voice-Steuerung zu erweitern, was der oben beschriebenen wechselseitigen Aufschaukelung Vorschub leistet.

9.7 Hinter jeder Suche steht ein Bedürfnis

Traditionelle Websites verlangen vom Nutzer, einen Großteil der Arbeit selbst zu erledigen. Die Betreiber sagen: Finde, was Du suchst und dann werde ich danach handeln und Wissen und Ware vorstellen. Ist eine Kundin auf der Suche nach ein paar Schuhen und einer passenden Handtasche, muss sie auf der Website zunächst nach Schuhen und dann nach Handtaschen suchen. Recommendation Engines können mehr oder weniger passende Vorschläge machen. Basierend auf Statistik: Kundinnen, die diese Schuhe gekauft haben, haben statistisch auch jene Handtaschen gekauft. Die statistische Grundlage derartiger Vorschläge ist aber häufig sehr instabil und die Vorschläge selbst sind daher entweder vollkommen generisch oder abwegig. Nutzer einer Website wollen wissen, wie Retouren funktionieren, oder Informationen über Versandkosten bekommen. Sie möchten erfahren, ob die Produkte aus umweltgerechter Erzeugung stammen. Das Grundinteraktionsmodell des heutigen Internets ist entweder sehen und klicken – oder die Suche findet, was man sucht. Häufig aber ist die Suche nicht der beste Weg herauszufinden, wie umweltfreundlich beispielsweise ein Produkt ist. Bei Conversational User Interfaces und intelligenten Assistenten ist das Interaktionsmodell grundlegend anders. Es ist viel schneller und direkter: „Besorge mir, was ich will“. Aus Nutzersicht sieht das so aus: „Ich will Schuhe und eine passende Handtasche kaufen“ und der Assistent erledigt die Aufgabe. Der Kunde muss keine Website öffnen und irgendwo versteckt finden, wie eine Retoure funktioniert, sondern sagt: „Ich will die Schuhe zurückgeben, die ich gerade geliefert bekommen habe“. Was für manchen wie ein marginaler Unterschied wirken mag, ist tatsächlich eine große Veränderung. Nicht Menschen müssen lernen, wie sie mit Maschinen umgehen können, sondern Maschinen müssen lernen, wie sie mit Menschen umgehen können. Das hat Auswirkungen auf die Entwickler der Maschinen und diejenigen Unternehmen, die sie beauftragen. Es verändert das gesamte Ökosystem.

Um das Kundenbedürfnis verstehen zu können, muss die Absicht in der Kundenaus sage identifiziert werden. Das ist im Fall eines Shopping-Assistenten in der Regel keine endlose Zahl verschiedener denkbarer Bedürfnisse und Absichten (Intents), sondern

eine endliche Anzahl von wenigen dutzend Fragestellungen. Gleches gilt auch für Banken, Reiseunternehmen, Telekommunikationsunternehmen – jede Branche hat eine begrenzte Anzahl solcher Fragestellungen, auf die sich Unternehmen künftig ausrichten müssen und die es ermöglichen, Kunden viel intensiver mit dem Versprechen zu umgarren: „Ich gebe Dir was Du willst!“. Der indisch-amerikanische Autor Kannan, P.V. beschreibt dies so: „Intents sind die Bausteine der Beziehung zwischen Kunden, Unternehmen und Institutionen“. Intent-Recognition ist daher eine der Kernaufgaben im Natural Language Processing und wird zu einer der Schlüsselfähigkeiten eines jeden Unternehmens werden.

9.8 Explizit und implizit

Von der Auflösung der eindeutigen Strukturen bei Smartphone-Apps, dem De-Bundling und Re-Bundling von Services und Inhalten war in diesem Kapitel bereits die Rede. Die Entsprechung der App im Feld der großen Assistenten sind bekanntlich die Skills und Actions, deren Zahl aktuell förmlich explodiert. Diese Actions und Skills entdeckt der Nutzer in der Regel in den zugehörigen Skill-Actions-Stores, ganz ähnlich wie im App-Store. Nachdem der Nutzer sich für einen dieser Skills entschieden hat, kann er ihn künftig mittels Invocation-Wort öffnen und sagen: „Alexa! Öffne xyz“. Daraufhin wechselt das Interface mehr oder weniger stark in die Verantwortung des Drittanbieters, der meistens die NLP-Fähigkeiten des Assistenzsystems nutzt und seine Inhalte und Angebote eigenständig und weitgehend abgekapselt anbietet – bis der Nutzer die Anwendung beendet oder den Smartspeaker durch erneutes Aussprechen einer Invocation-Phrase sogar aktiv und explizit auffordert, einen bestimmten Dienst zu öffnen.

Neben dieser expliziten Form, gibt es eine implizite Form, wie Nutzer auf Inhalte und Dienste zugreifen. Diese implizite Form ist aus verschiedenen Gründen viel bedeutender (siehe Abb. 9.5). Einerseits, weil viele Nutzer das explizite Invocation-Word nicht kennen oder nicht erinnern, was bei einer täglich wachsenden Zahl von Actions und Skills sicherlich zukünftig nicht besser werden wird. Das Grunddilemma der App-Stores mit begrenztem Raum der Präsentation wird mit tausenden Angeboten in den Skill-Stores eher noch verschlimmert, ohne die Visualität als Merkstütze zur Verfügung zu haben.

Anstatt zu sagen: „Alexa! Öffne TV-Guide xyz“ und in der Folge zu fragen: „TV-Guide, was läuft heute um 20:15 im TV?“, ist eine implizite Invocation direkt. Eine solche implizite Invocation lautet: „Alexa, was läuft heute um 20:15 im Fernsehen?“. Implizite Invocations sind viel näher an einer Voice-Suchanfrage. Der Assistent wird nun daher parallel nach Skill-Anbietern suchen, die unter Umständen diesen Nutzer-Intent (TV-Programm 20:15 Deutschland) enthält und parallel eine Websuche starten. Im Fall, dass der Assistent passende Inhalte in einem Skill findet, wird die Antwort lauten: „Heute Abend läuft im Ersten der Tatort um 20:15. Weitere Informationen findest Du im TV-Guide XZY. Soll ich diesen für Dich öffnen?“, um dann direkt tief in den entsprechenden Skill zu springen und dort die Inhaltsbeschreibung des Tatorts vorzustellen.

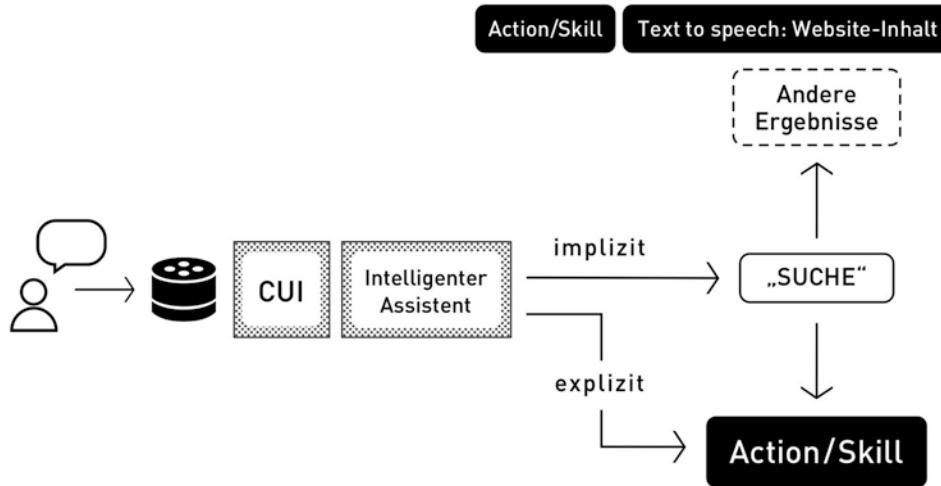


Abb. 9.5 Implizite und explizite Suche bei einem intelligenten Assistenten

Das Beispiel verdeutlicht, wie sich die Disintegration im CUI-Ökosystem fortsetzt, die im App-Store bereits im Gange ist. Im Fall von CUI passiert dies nicht erst nach einer langen Marktentwicklungsphase, sondern direkt zu Beginn. In beiden Fällen verlieren Anbieter die Kontrolle über die User Journey. Um in diesem Markt bestehen zu können, ist es wichtig, beides zu betreiben. Ein eigenes Angebot zu unterhalten und offen für die Angebote anderer zu sein. Wie die Suchmaschinenbetreiber heute Guidelines veröffentlichen, auf welche Weise Inhalte aufbereitet sein sollen, damit sie von den Suchmaschinen optimal erfasst werden, entwickeln sich derzeit auch Guidelines, die für die Fragen rund um Assistenten und Voice Search bedeutend sind. Sobald der Nutzer auf eine vom Anbieter kontrollierte User Journey einschwenkt, beispielsweise weil ein Produkt gekauft wurde, muss der Kunde durch überragenden Service, Einfachheit in der Abwicklung und Erkennen seiner Bedürfnisse für zukünftige Transaktionen gehalten werden.

In diesem Denken ist Voice Search Optimization und der Betrieb eigener Skills/Actions, sowie das Bemühen, auf vielen Plattformen, die im Rahmen einer Websuche durch den Assistenten zu Rate gezogen werden, optimal auszugestalten. Bei lokalen Anbietern wird hier beispielsweise auf die Bedeutung eines zutreffenden Yelp-Profil oder die Bedeutung der Qualität des Wikipedia-Eintrags verwiesen. Wie gut positioniert internationale Marken sowie nationale und lokale Akteure sind, wird inzwischen regelmäßig in Studien erfasst und gebenchmarkt. Die Frage, wie gut eine lokale Orthopädie-Praxis durch Assistenten aufgefunden wird, kann in nicht allzu ferner Zukunft wesentlich für den Erfolg dieses kleinen Unternehmens sein. Was geschieht, wenn die Frage lautet: „Was ist Coca Cola?“ Aus Unternehmenssicht sollte die Antwort nicht lauten: „Coca Cola ist ein Wettbewerber von Pepsi“ Jeder Händler möchte auf die Frage: „Wo kann ich eine Damendaunenjacke kaufen?“ als erster Anbieter gelistet sein. All

dies sind Aufgaben, die technische, inhaltliche und unternehmerische Fragen betreffen und daher so schwierig zu lösen sind, wie viele vorherige Fragen der Digitalisierung. Nur ist diesmal die Komplexität deutlich höher, da das relevante Universum viel breiter ist.

9.9 Auffinden von Services und Produkten

Wie kann es Marken und Anbietern in diesem System gelingen, die Aufmerksamkeit auf ihre Angebote zu lenken? Welche Rolle spielt Werbung in diesem Gefüge? Werbung ist im Internet die mit Abstand wichtigste Refinanzierungsquelle und wesentliches Element im bestehenden Ökosystem. Werbung wurde in diesem Zusammenhang schon mehr als einmal totgesagt – zuletzt, als die Frage aufkam, wie Werbung auf den kleinen Smartphone-Bildschirmen überhaupt von Nutzern akzeptiert wird. Tatsächlich ist die Effizienz von mobilen Werbeformen trotz der wesentlich besseren Möglichkeit des Targetings nicht sehr hoch. Im sich jetzt entwickelnden Ökosystem ist durchaus denkbar, dass Werbung – im Sinne einer begleitenden, störenden Zugabe zu Content und/oder Funktionen gar nicht mehr stattfinden wird. Schon deshalb, weil in einer dialogischen Welt kein Platz für Banner am Bildschirmrand ist. Die Effizienz von Werbung wird angesichts des vollständig personalisierten Nutzererlebnisses ohnehin schrecklich schlecht sein, weil der Assistent ja alles von mir weiß. Den Kontext in dem sich der Nutzer befindet (wo, woher, wohin, warum ...) erfasst und darauf abgestimmt agiert. Der Assistent überprüft selbsttätig, welche Angebote es gibt – Werbung heißt daher möglicherweise künftig eher API-Marketing. Also eine Bewerbung bei der Maschine und nicht beim Nutzer. Der intelligente Assistent ist der perfekte Adblocker. Selbst das sonst in Bezug auf Monetarisierung der Nutzer-Aufmerksamkeit nicht besonders zimperliche Facebook weiß, dass Zurückhaltung im Conversational Raum wichtig ist. Chatbots, die auf Facebook Messenger angeboten werden, können mit Nutzern nur Kontakt aufnehmen, wenn der Nutzer zuvor aktiv den Chatbot angesprochen hat. Es ist einfach nicht vorstellbar, dass der Smartspeaker beginnt, einem die aktuellen Angebote des Supermarktes um die Ecke vorzulesen, sobald man die Wohnungstür beim Nach Hausekommen hinter sich geschlossen hat. Etwas subtilere Formen sind möglicherweise schon denkbar:

- Nutzer: „Hallo Google, was sind die Neuigkeiten von der NASA?“
- Google Home: „Wissenschaftler bei der NASA haben kürzlich ein neues Sternensystem im xyz entdeckt. Das meldet der Service example.com. Ich habe Links zu zwei Artikeln an dein Smartphone geschickt. Soll ich dir den ersten jetzt vorlesen? Jeder Artikel kostet 10 Cent“

Wie schmal der Grat von nützlich zu unerträglich sein kann, erkennt man auch an diesem Beispiel [6]: Google wird vom Nutzer nach dem Wetter für den heutigen Tag gefragt und nachdem der Assistent die Werte aufgezählt hat, unterbricht eine radiower-

Abb. 9.6 Google Home liefert unerwünschte Werbung während des Wetterberichts (Foto: Twitter)



bungähnliche Einlage, in der schallend ein Angebot zum Disney-Vergnügungspark am kommenden Wochenende unterbreitet wird (siehe Abb. 9.6). Der Nutzer kommentiert richtig auf Twitter: „Was noch niemand je gesagt hat: Ich wollte schon immer eine Uhr mit Werbung haben“

Es darf allerdings nicht der Eindruck entstehen, dass Marken und Marketingkommunikation im Zeitalter von Assistenten ausgedient hätten. Zunächst werden natürlich all die Kunden-Touchpoints an Bedeutung gewinnen, die noch nicht über Geräte wahrgenommen werden, wie alle Formen von Out-of-Home-Kommunikation. Auch die Art, wie sich Marken im digitalen Raum der Assistenten und CUIs darstellen, wird eine größere Bedeutung bekommen. Sonic Branding war bislang eher die Frage einer markanten Tonfolge am Ende eines TV-Werbespots. Als hörbares Branding im Zeitalter von Voice und Assistenten wird es ein ganz zentrales Werkzeug. Spricht die „Allerweltsstimme“ von Alexa, eine eigene synthetische Stimme oder leistet man sich einen Voice Actor, der das eigene CUI repräsentiert? Ist dies womöglich die Stimme des CEO des Unternehmens? Man kann Stimmen synthetisieren und damit vermeiden, dass der viel beschäftigte Manager tausende von möglichen Antworten einsprechen muss. Was aber, wenn der CEO wechselt?

9.10 Findability und Discoverability

Auf die Fragen nach der Auffindbarkeit (Findability) und nach den Möglichkeiten, Dinge zu entdecken (Discoverability) sind bislang im Feld von CUIs und intelligenten Assistenten nur wenig befriedigende Antworten gefunden worden. Betrachtet man das Geschäft von Radioanbietern, fällt auf, dass diese vom Trend der Nutzerabkehr von linearen Formaten noch erstaunlich unberührt sind. Während Fernsehsender Probleme bekommen, weil Streamingdienste jederzeit die gewünschte Unterhaltung bieten und es auch verstehen, den Nutzer durch entsprechende Empfehlungen an den Service zu binden, haben Radiostationen trotz Musikstreaming ihren Platz gut behaupten können. Nicht zuletzt weil sie als typisches Nebenbei-Medium beim Autofahren oder sonstigen Tätigkeiten bequem genutzt werden. Sie sind in der Auswahl von Inhalten im Bereich News oder Entertainment in ihrer Nutzungssituation akzeptiert. Durch die starke Entwicklung im Bereich non-linearer Audioformate wie sie Podcasts darstellen, kommt nun aber doch Bewegung ins Feld der Anbieter von Audioinhalten. Natürlich produzieren auch Radio-Macher Podcasts. Das Problem bleibt aber die schlechte Auffindbarkeit. Die wenigsten Podcasts und Radiobeiträge traditioneller Sender liegen in einer maschinenlesbaren Form vor. Die Transkription ist komplexer als vielleicht zunächst erwartet, denn das weite Feld der unterschiedlichen Themen und Wissensdomänen und die heterogene Art der Audioinhalte – verschiedene Stimmen, die sich teilweise ins Wort fallen, die Slang und Umgangssprache, Atmosphäre und O-Töne machen eine automatisierte Transkription sehr anspruchsvoll.

Ohne Transkripte aber können Audioinhalte nicht gesucht, gefunden und im Zusammenhang mit Sprachassistenten verwendet werden. Google hat eine Initiative gestartet, die Podcasts identifiziert, um damit eine „In-Episoden-Suche“ [7] zu ermöglichen, also das spezifische Auffinden bestimmter inhaltlicher Stellen in einem längeren Hörstück (also auch hier gibt es ein De-Bundling und Re-Bundling...). Diese Initiative steht aber offensichtlich noch am Anfang. Viel weiter scheint das Unternehmen Audioburst [8] aus Israel zu sein. Es hat sich als Audio-Personalisierungsspezialist positioniert. Es ist eine Findmaschine für Audio-Content, um beim Fahren, beim Kochen, beim Sport und Reisen jeweils gezielten Zugriff auf die gewünschten Audioinhalte zu bieten. Grundlage ist ein gut ausgebautes NLP mit offensichtlich beeindruckenden Transcription-Fähigkeiten. Die öffentlichen amerikanischen Radiosender NPR veröffentlichen bereits seit mehreren Jahren eine repräsentative Studie, in der die Nutzungsarten von Smartspeakern erfasst werden. Jedes Jahr wird daraus ersichtlich, dass die meisten Nutzer die Smartspeaker anstelle eines Radios in der Küche, am Arbeitsplatz, oder im Auto benutzen.

Viele Radioanbieter in USA und Deutschland haben allerdings versäumt, rechtzeitig eine Strategie des Zangenangriffs aus Podcast-Welle, Streamingdiensten und Smartspeakern zu finden. Nun müssen sie erleben, wie ihre linearen Radioprogramme von Aggregatoren in Alexa abgespielt werden und nicht von ihren eigenen Skills. Und sie er-

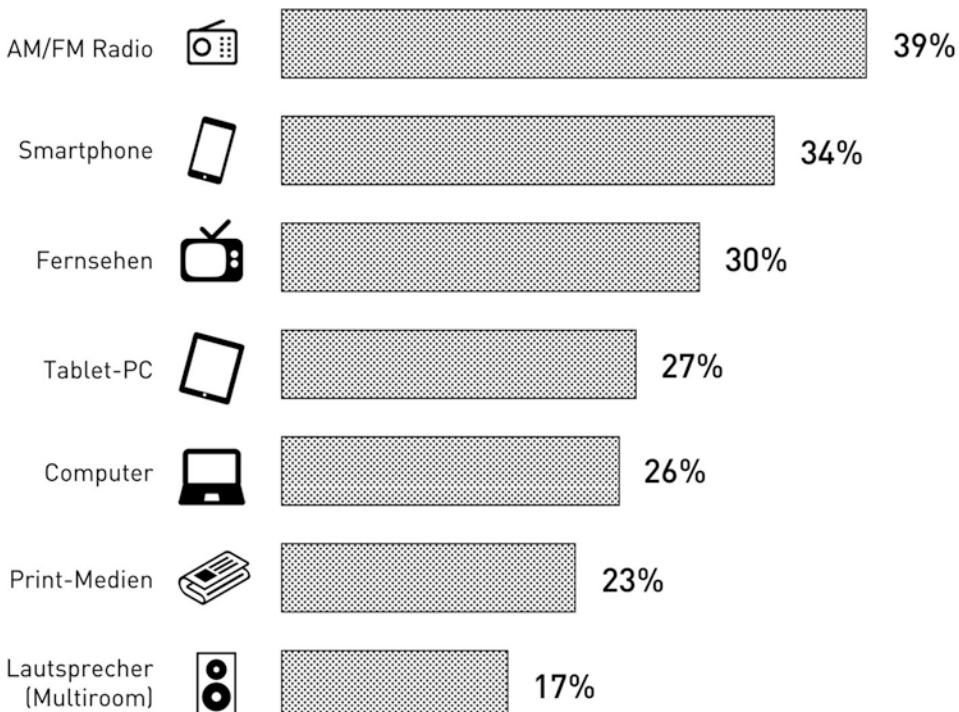


Abb. 9.7 CUIs gehen zeitlich zu Lasten der Mediennutzung von

leben, wie ihre Audio-Inhalte im Tsunami der Podcasts aufgrund fehlender Transkripte und Metainformationen nicht aufzufinden sind und die Veränderung des Ökosystems ihre ganze Welt zerrüttet, die die Digitalisierung bislang relativ unbeschadet überstanden hatte (siehe Abb. 9.7).

9.11 Customer Journey Commerce

Nicht nur Anbieter von Inhalten müssen ihren Weg zum Nutzer und zu neuen Kunden finden und dabei eine Abhängigkeit von großen und kleineren Aggregatoren vermeiden. Auch und vor allem Anbietern von Waren und Dienstleistungen geht es genauso. In diesem neuen Ökosystem gibt es neue Customer Journeys. Sie unterscheiden sich vor allem darin, ob es sich bei dem Produkt um ein Low-Interest-Produkt handelt, wie es wohl die meisten Packaged Goods sind, oder um eine Produktkategorie, um die sich der Nutzer bewusst kümmert, wie zum Beispiel die Auswahl eines Restaurants. Oder ob es sogar ein High-Interest-Produkt ist, also Mode, Gesundheit oder die Fernreise mit Kulturprogramm. Weiterhin unterscheiden sich die Customer Journeys teilweise danach, ob der Nutzer

bereits eine Produktpräferenz entwickelt bzw. geäußert hat. Die Nachbestellung von Toilettenpapier verläuft naturgemäß anders, als die Bestellung eines neuen Outfits. Je niedriger das Interesse und je ausgeprägter die bereits entwickelte Produktpräferenz, desto autonomer kann der Assistent eine Entscheidung treffen. Umso kürzer fällt die Customer Journey aus. Vertraut der Nutzer dem Assistenten, folgt auf die Äußerung des Bedürfnisses unmittelbar eine Bestellung. Da ist kein Platz für Werbung und wenig Spielraum für die Manipulation seitens der Anbieter. Es ist ein interessanter Ort für Marken mit Gattungsbedeutung (z. B. Tempo, Tesa, Edding). Wir wissen allerdings, wie wenigen Marken eine derartige Stellung im Sprachgebrauch gegückt ist.

9.12 Berührungsloser Konsum

Wenn die Bestellung lautet: „Alexa kaufe eine Packung Tempo-Taschentücher“, kann die Antwort nicht lauten: „Gerne, aber wir haben derzeit Amazon-Prime-Tücher im Angebot“. Im Fall, dass keine Präferenz festgelegt ist, kann der Assistent natürlich einige Vorschläge unterbreiten. Aus wahrnehmungpsychologischen Gründen liegt die Anzahl maximal bei drei und bietet Raum für eine Empfehlung seitens des Assistenten. Es gibt unter Umständen Raum für weitere Produktspezifikationen über den Preis hinaus. Diese zu definieren und dabei das oben genannte nutzerzentrierte, semantische Prinzip anzuwenden, ist entscheidend, wenn Anbieter in dieser Auswahl auftauchen wollen. Wenig überraschend gibt es bei Amazon Versuche, die Platzierung in diesen wenigen Ergebnistreffern an Anbieter zu verkaufen. Ebensowenig überraschend erwarten viele Marktteilnehmer, dass Amazon seine eigenen Prime-Produkte hier möglicherweise auch wettbewerbswidrig priorisieren wird. Das allerdings wird im Einzelnen ähnlich schwer nachzuweisen sein, wie es bereits der Fall ist bei Google-Suchergebnissen oder den Rankings der vielen Energie- und Telefontarif-Agggregatoren.

Im Rahmen einer international durchgeföhrten qualitativen Studie [9] hat Ericsson Ende 2018 über 30 Millionen Menschen online befragt. Die Hälfte der befragten Nutzer von virtuellen Assistenten möchte die Wiederaufstockung von Alltagsbedarf im Haushalt automatisieren. Sie wollen zudem den Überblick über die Rechnungen für Gas, Strom und Wasser durch Bots behalten und gaben sogar den Wunsch an, dass ein virtueller Assistent alle ihre Finanzen, einschließlich der Steuererklärungen verwaltet. Die gleiche Anzahl der Befragten wünscht, dass Assistenten Garantiereparaturen der Haushaltselektronik automatisiert verwalten. Ein derartiger berührungsloser Konsum, bei dem Toilettenpapier und Versicherungspolicen automatisiert bestellt werden, ist eine gravierende Veränderung für das Marketing vieler Unternehmen.

Für den Fall, dass der Nutzer bei derartigen eher wenig interessanten Produktkategorien mit den zunächst gemachten Vorschlägen nicht einverstanden ist, wird der Assistent voraussichtlich das Medium wechseln und vorschlagen, eine komplette Liste mit Wahlmöglichkeiten im Messenger-Window oder per E-Mail zu verschicken. Diese multimodale Verteilung der Customer Journey wird wohl der Regelfall bei High-Interest-Angeboten werden. Die hochauflösenden Bilder des karibischen Sandstrands wird der Assistent vor-

schlagen, auf dem heimischen Großbildschirm zu betrachten und die Auswahl zwischen drei Handtaschen möglicherweise sogar in Form einer AR-Anwendung in einer virtuellen Anprobe.

Wenn Discovery (Auffindbarkeit) ein bislang zumindest ungelöstes Problem darstellt und wir daher immer stärker darauf angewiesen sind, den Empfehlungen von Maschinen zu trauen, stellen sich Fragen: Was mache ich, wenn ich gar nicht weiß, was ich möchte? Wonach kann ich fragen? Welcher Algorithmus ist nach meinem Geschmack? Ist mein Geschmack womöglich ein Algorithmus? Wie wertvoll Vertrauen in die Anbieter in dem neuen Ökosystem ist, haben wir schon mehrfach betont.

9.13 Mentale Adipositas

Die Automatisierung von Entscheidungen kann uns mental faul machen. Toast oder Müsli? Was wäre, wenn Sie sich nicht einmal entscheiden könnten, was Sie zum Frühstück essen möchten, ohne ein digitales Gerät um Rat zu fragen? Dies könnte Realität werden, wenn man bedenkt, dass 34 Prozent der von Ericsson befragten Nutzer von intelligenten Assistenten einer KI als Ernährungsberater vertrauen würden, aber nur 25 Prozent einem Menschen. Die Automatisierung solcher Entscheidungen könnte Sie jedoch auf einen Weg führen, auf dem Sie nicht mehr wissen, wie Sie sich fühlen. In der Tat, 57 Prozent der Befragten der Ericsson-Studie sagen, sie wollen ein Smartphone, das weiß, wann sie krank werden, bevor sie es selbst bemerken. 54 Prozent möchten eine Smartwatch, die spürt, wann sie gestresst sind, bevor sie es bemerken. Würde dies dazu führen, dass Menschen an medizinischen Terminen teilnehmen, ohne zu wissen, warum? Dreißig Prozent wünschen sich einen intelligenten Assistenten, der entscheidet, wann er einen Arzt, Zahnarzt oder Friseur aufsucht. Das heutige Bildschirmzeitalter hat uns faul gemacht, da wir uns nicht bewegen müssen. Alles, was wir wollen, ist nur ein Klick oder ein Zug, und viele von uns trainieren unseren Körper weniger als notwendig. Aber wir stehen vielleicht an der Schwelle zu einer mentalen Faulheit, die sich auf unsere Gedanken ausbreitet. Mit der digitalen Entscheidungsunterstützung in immer mehr Bereichen des täglichen Lebens glaubt ein Drittel der befragten Verbraucher, dass intelligente Assistenten die Menschen vergessen lassen, wie sie ihre eigenen Entscheidungen treffen können. Tatsächlich glauben 34 Prozent bereits, dass kritisches Denken durch den übermäßigen Einsatz von intelligenten Assistenten verschwinden wird. 31 Prozent der Verbraucher erwarten, dass wir in „Mind Gyms“ gehen müssen, um das Denken wieder zu trainieren.

9.14 Walled Gardens vs. offene Architekturen

Sind nun Drittanbieter und Nutzer künftig auf Gedeih und Verderb abhängig von einer ganz kleinen Anzahl vollintegrierter Plattformen wie sie Google, Facebook und Tencent anbieten? Bewegen wir uns in eine Zeit, in der die Welt in so genannte Walled Gardens aufgeteilt ist, wie dies einst AOL oder Compuserve darstellten, die einen Zugang zum

freien Internet anbieten und mit einem eigenen Layer von Funktionalitäten ausstatteten? Schaut man auf die darauf folgende Evolutionsstufe des Mobile-Ökosystems, in der ebenfalls strikte Walled Gardens aufgebaut wurden und bis heute existieren, kann man vermuten, dass dies auch die gegebene Struktur des CUI-Ökosystems sein wird. Für einen solchen Verlauf spricht auch, dass derjenige, der viele Daten sammelt, seine Services schneller weiterentwickeln und verbessern kann, und dass das Betreiben eines vollintegrierten Tech Stack von NLP bis zur Erfüllung von Kundenwünschen in geschlossenen Systemen häufiger besser gelingt, als in offenen. Wie beschrieben ist Apple mit seiner starken Abgrenzung und Kontrolle ein gutes Beispiel für die Funktionalität eines solch geschlossenen Systems.

Gegen eine solche monolithische Entwicklung spricht, dass die Aufgabe ungewöhnlich groß ist. Wie am Anfang des Kapitels beschrieben, erweitert sich das relevante Universum gerade in jede denkbare Dimension und spannt einen Bogen von Home-Automation über Industrieautomation bis zu Sprache und Logistik. Bei allem Respekt für die Dimensionen, in die sich Unternehmen, wie Google und Amazon entwickelt haben und die Ambitionen, die beispielsweise in China von Unternehmen und Staat ausgeht, gibt es kein Beispiel auf der Welt, das eine derartige Komplexität unter einem Dach darstellen könnte. Vor diesem Hintergrund halten wir es nicht nur für nicht wünschenswert, sondern auch für wenig wahrscheinlich und gehen davon aus, dass Kernfunktionalitäten in gewisser Weise zur Commodity werden und mit eigenen Datenpools und Verbindungen verwendet werden. Wie in Kap. 8 deutlich wird, können Unternehmen NLP-Funktionalitäten aus dem Netz beziehen und mit eigenen Kundendaten, weiteren Third-Party-Daten, Wetter und ähnlichem kombinieren und einen eigenen Assistenten-Service aufbauen und betreiben, den man als Skill auf Amazon Alexa und zugleich auf der eigenen Website sowie der eigenen Mobile-App Nutzern anbieten kann. Ein heterogenes System, bei dem Anbieter aus unterschiedlichen Bereichen Beiträge leisten. Ikea bietet Lampen an, die mit Amazon Alexa steuerbar sind, Autohersteller wie BMW oder Ford integrieren Alexa-Funktionalitäten und gleichzeitig Google Android for Car. Auf der jährlichen Computerelektronik-Messe CES Anfang 2019 kämpften alle Anbieter um den Spaltenplatz, wer mehr Lizenzpartner präsentieren konnte und wessen Smart-Assistant-Integration am verwegsten war. Amazon gab die Lösung aus „Alexa Everywhere“ [10]. Unternehmen dagegen müssen sich nicht entschieden. So bietet GE Lampen an, die sowohl mit Alexa als auch mit Google gesteuert werden können.

Auch auf einer anderen Ebene des Ökosystems sind die Grenzen der Systeme bereits heute teilweise durchlässiger, als es erscheint. Microsoft integriert seinen Assistenten in Alexa und ermöglicht damit Office365-Funktionen über den Smartspeaker zu steuern, beispielsweise einen Kalendereintrag zu verfassen und Apples Siri öffnet Nutzern auf Sprachbefehl auch den Google Assistant und lässt darüber die Google-Funktionen wie Maps und vieles andere mehr über die Plattformgrenzen hinaus nutzen.

Es fällt nicht schwer, sich eine Zukunft vorzustellen, in der Nutzer aus einer ganzen Reihe von CUIs und intelligenten Assistenten auswählen können und diese auf ihren Geräten – vom Smartphone über das Auto bis zum Kühlschrank – installieren. Wie ein eige-

ner Klingelton zeitweilig ein Ausdruck der eigenen Persönlichkeit war, so kann die Wahl eines Assistenten die aktuelle Stimmung ausdrücken. In einem voraussichtlich offenen System liegt darin natürlich eine große Chance für Unternehmen und Institutionen aller Art, derart eigenständige Angebote zu machen.

So wünschenswert es für Unternehmen und Institutionen ist, eigene CUIs zu entwickeln, die auf verschiedenen Plattformen funktionieren, über verschiedene Kanäle vertrieben werden können und auf sehr unterschiedlichen Endgeräten genutzt werden können, so komplex ist diese Aufgabe in der Realität. Wer sich jemals mit so genannten responsiven Website-Konzepten beschäftigt hat, also Websites, die auf einem Desktop-Webbrowser genauso optimiert sind, wie für ein Tablet, das keine Mausbefehle, dafür aber Touch-Bedienung kennt, und die gleiche Website auf einem vertikal ausgerichteten Smartphone-Bildschirm funktioniert, weiß, wie kompliziert das sein kann. Es gilt immer eine Entscheidung hinsichtlich der Funktionen zu treffen: Desktop first? Mobile first? Das Ergebnis ist nicht selten der kleinste gemeinsame Nenner, der keinen Anwendungsfall optimal bedient.

Ähnlich, nur ungleich komplexer, ist es, Systeme aufzusetzen, die für Sprache und Text gleichermaßen gut geeignet sind. Ein Problem ist bereits die zeitliche Dimension: synchrone vs. asynchrone Kommunikation. Im Sprachmodus muss ein Dialog jeweils unmittelbar fortgesetzt werden oder er wird abgebrochen und beendet. Im Messenger kann zwischen einer Antwort und einer Wiederaufnahme des Dialogs eine unbestimmte Zeit liegen, was den Dialogfluss erheblich beeinflusst. Während es im Messenger von Chatbots Möglichkeiten für grafische Bedienelemente und visuelle Informationen gibt, steht dies bei Voice-Only oder sogenannte „Busy-Eye-Modes“ – beispielsweise während einer Autofahrt – natürlich nicht zur Verfügung. Nahezu jede Bot-Builder-Plattform, ob groß oder klein, verspricht, dass Bots, einmal auf der betreffenden Technikplattform gebaut, über alle Kanäle genutzt werden können. Die Realität sieht natürlich ganz anders aus. Die Interfaces sind verschieden und müssen entsprechend jeweils einzeln betrachtet werden. Dennoch erwarten die Nutzer eine nahtlose Verzahnung dieser multimodalen Ebenen – und wer im Ökosystem bestehen möchte, muss sich damit auseinandersetzen.

9.15 Transaktionen

CUIs und intelligente Assistenten sind Bedürfniserfüllungsgehilfen. Sehr viele menschliche Bedürfnisse münden automatisch in der Notwendigkeit einer Transaktion, eines Kaufes von Gütern oder Diensten. Anbieter, die Transaktionen ermöglichen, entwickeln eine Stickiness bei Nutzern. Daher sind neben Nutzerpräferenzen die damit verbundenen Finanztransaktionselemente eine wesentliche Säule des Ökosystems. Marketers weltweit sind bezüglich Voice-Shopping höchst optimistisch. In einer Anfang 2019 von Voice.com in Zusammenarbeit mit Voicebot.ai durchgeföhrten Befragung [11] amerikanischer Entscheider äußerte sich deutlich über die Hälfte der Befragten als sehr optimistisch. Selbst wenn die aktuellen Zahlen von Amazon und Co dies zumindest im westlichen Teil der

Welt noch nicht bestätigen. Anders ist die Situation in Indien (siehe Niki.ai Kap. 2) und China, wo WeChat eine Conversational Plattform ist, die durch die Fähigkeit zu bezahlen besonders schnell gewachsen ist und Alibaba über seinen TMall Genie Assistenten dem Vernehmen nach sehr erfolgreich verkauft. All diese Anbieter können dies nur tun, da sie die Möglichkeit zu bezahlen in ihren Systemen integriert haben. Da bei Amazon nur registrierte Nutzer Alexa nutzen können und die Hinterlegung eines geeigneten Zahlungsmittels, zum Beispiel einer Kreditkarte, Voraussetzung für die Durchführung eines Einkaufes ist, liegt die Latte recht hoch. Ähnlich agieren auch andere E-Commerce-Anbieter, die mit Voice Commerce experimentieren. Otto.de bietet die Möglichkeit, ausgewählte Angebote über den Google Assistant zu kaufen, was voraussetzt, dass man sich zuvor auf Otto.de registriert hat und das Google-Konto mit dem Otto-Konto verknüpft. Bei Niki.ai ist der Weg anders. Niki.ai ist ein Dienst, der vor allem in Mobile-Banking-Apps integriert ist und damit automatisch auf die gewählte Kontoverbindung zugreift, wenn der Nutzer etwas über Niki.ai kauft. Dieser Weg zum Nutzer ist in Asien recht weit verbreitet, da Mobile-Wallet-Anbieter, wie beispielsweise PayTM in Indien, enorme Marktanteile haben und ihre Apps durch intelligente Assistenten ergänzen, deren Transaktionen auch das eigene Transaktionsvolumen steigern. Die rege Aktivität von Apple Pay als ein international funktionales Paymentsystem zu installieren und gleiche Vorstöße von Google (Google Pay) zeigen, dass auch diese Unternehmen verstehen, wie entscheidend die Zahlungsmöglichkeiten und die Verknüpfungen mit den jeweiligen Assistenten sind. Es wird interessant sein zu beobachten, wie Facebook sein kontrovers aufgenommenes Payment Konzept Libra in den Markt bringen kann, und ob dies dem Unternehmen möglicherweise den dringend benötigten Schub geben kann, im CUI- und Assistenten-Markt als Spätstarter noch aufzuholen.

Wenn es um Payment geht, stellt sich die Frage nach der Authentifizierung. Menschliche Stimmen haben genau so ein individuelles Profil wie Fingerabdrücke und es gibt verschiedene Anwendungen von biometrischen Authentifizierungssystemen, die ausschließlich über Stimme funktionieren. Allerdings steht die Integration derartiger Dienste noch ganz am Anfang. Amazon ermöglicht es theoretisch, mehrere Accounts auf einer Alexa zu betreiben, was in der Praxis allerdings nicht befriedigend funktioniert. Da biometrische Authentifizierung über Fingerabdruck und Gesichtserkennung bereits Bestandteil der höherwertigen Smartphones sind, stellt sich die Frage, ob die Authentifizierung künftig nicht eher auf einem Gerät stattfinden wird, anstelle der Cloud, wie dies nicht zuletzt aus Datensicherheitsgründen und Performancegründen im Bereich von NLP zunehmend stattfindet.

Literatur

1. Kinsella, Bret: Adam Cheyer of Samsung and Viv Labs Talks 25 Years of Voice Assistants – Voicebot Podcast Ep 109, (13.08.2019), <https://voicebot.ai/2019/08/13/adam-cheyer-of-samsung-and-viv-labs-talks-25-years-of-voice-assistants-voicebot-podcast-ep-109/>, zugegriffen 12.11.2019

2. Brown, Ben: How Apple Stock Reacts After Every Major iPhone Reveal, (10.09.2019) <https://www.cnn.com/apple-stock-iphone-launch/>, zugegriffen 10.11.2019
3. YouTube: Apple Knowledge Navigator (2017), <https://youtu.be/p1goCh3Qd7M>, zugegriffen 22.11.2019
4. Kittlaus, Dag: Beyond Siri – The World Premiere of Viv with Dag Kittlaus, (09.05.2016), <https://techcrunch.com/video/beyond-siri-the-world-premiere-of-viv-with-dag-kittlaus/>
5. Samsung: The bixby Marketplace is now officially open in the U.S. and Korea, (01.07.2019), <https://news.samsung.com/global/the-bixby-marketplace-is-now-officially-open-in-the-us-and-korea>, zugegriffen 30.10.2019
6. Meunier, Bryson: New Beauty & the Beast promo is one way Google could monetize Home, Twitter, 16.03.2018, <https://twitter.com/i/status/842358950536318976>
7. Wang, Jules: Google Podcasts in-episode search is coming, shows now being fully transcribed, (26.03.2019), <https://www.androidpolice.com/2019/03/26/google-podcasts-in-episode-search-is-coming-shows-now-being-fully-transcribed/>, zugegriffen 20.11.2019
8. Audioburst, <https://www.audioburst.com/>
9. Ericsson: Ericsson releases its 10 hot consumer trends for 2019, (12.12.2018), <https://www.ericsson.com/en/press-releases/2018/12/ericsson-releases-its-10-hot-consumer-trends-for-2019>, zugegriffen 20.11.2019
10. Vavrek, Adam: Amazon Alexa at CES 2019, (26.12.2018), <https://developer.amazon.com/de/blogs/alexa/post/5dfc80ef-7f2c-44ab-82f2-45d79cd1394f/amazon-alexa-at-ces-2019>, zugegriffen 03.11.2019
11. Schwartz, Eric Hal: Report: 31M Americans Will Shop by Smart Speaker in 2019, (10.07.2019), <https://voicebot.ai/2019/07/10/report-31m-americans-will-shop-by-smart-speaker-in-2019/>, zugegriffen 10.11.2019



Reality Check

10

Zusammenfassung

In den vorherigen Kapiteln ging es vor allem um das, was noch kommen wird. Doch wo stehen wir heute? Der Reality Check zeigt die Hürden und Hindernisse auf dem Weg zum perfekten CUI. Hier geht es um stockende Dialoge, fehlende Emotionen, mangelnde Mehrsprachigkeit und die (noch sehr) groben Werkzeuge, mit denen CUI-Entwickler arbeiten müssen. Multimodalität ist ein wichtiges Stichwort: Nur wenn Auswahlprozesse ohne Brüche in der Nutzung Produkte auf einem Bildschirm anzeigen, werden CUIs Erfolg haben. Letztendlich bleiben noch viele offene Fragen zum Thema Datenschutz und Datensicherheit. Oder wissen Sie bereits, wie man Adversarial Attacks in den Griff bekommt?

Wir haben verschiedene Visionen und Benchmarks vorgestellt. Von Google Duplex des Jahres 2019 über das ursprüngliche Siri aus dem Jahr 2010 und Viv.ai aus 2016, das heute seinen Dienst im Samsung Bixby versieht, bis hin zur übergroßen Apple Knowledge Navigator Vision aus dem Jahre 1987. Doch wo stehen wir in der Entwicklung heute tatsächlich? Die meisten Nutzer dürften bereits erste Erfahrungen mit Siri oder Alexa, einem Chatbot oder dem Bot einer Servicehotline gemacht haben.

Ist die Realität nun eher bei Duplex, der fast menschliche Dialoge führt oder bei Clippy, der animierten Büroklammer von Microsoft, die 1997 – zunächst nicht deaktivierbar – Bestandteil des Office-Pakets wurde. Sie war eine der ersten virtuellen Assistenten, mit der jeder Nutzer zwangsläufig in Berührung kam oder besser: Sie war der Versuch eines Assistenten. Clippy sah zwar niedlich aus, war aber nur selten beim Bearbeiten eines Word-Dokuments tatsächlich hilfreich. Diese Funktion schien sich in all den Jahren mit diversen Word-Versionen nie weiter zu entwickeln.

Das ist bei CUIs anders. Während Googles Duplex noch Zukunftsmusik ist, entwickelt der Konzern die Technologie weiter und bietet bereits erste kommerzielle Produkte an. Ein Beispiel ist CallJoy [1]. Zielgruppe für den Telefonassistenten sind die über 30 Millionen kleinen und mittelständischen Unternehmen, die sich kein klassisches Call-Center leisten können. Oder aber Einzelhändler, die am Telefon von persönlichen Beratungsgesprächen in ihrem Ladengeschäft abgehalten werden. Der Dienst kostet 39 Dollar im Monat und filtert beispielsweise Spam-Anrufe. Kernaufgabe ist allerdings die Beantwortung der FAQ: Wie lange hat das Geschäft geöffnet? Kann ich vor dem Laden parken? Ist Produkt xy vorrätig?

Möchte man ein Produkt bestellen oder einen Termin vereinbaren, fragt der Assistent: „Darf ich Ihnen einen Link per SMS senden?“ Stimmt der Nutzer zu, erhält er oder sie einen Link, um den Bestellvorgang oder die Terminvereinbarung abzuschließen.

Nur bei Fragen, die CallJoy nicht beantworten kann, wird auf die normale Rufnummer des Geschäfts durchgestellt. Im Idealfall verbessert der Inhaber mit dem Assistenten seinen telefonischen Service und verschafft sich gleichzeitig mehr Freiraum, um für seine Kunden im Laden verfügbar zu sein. Wie wir in diesem Buch in diversen Beispielen darlegen, sind international an vielen Stellen sprachbasierte Systeme erfolgreich live. Wie wir aber ebenfalls beschreiben, steht die Entwicklung noch am Anfang – und damit verbunden sind viele enttäuschte Erwartungen auf Seiten von Betreibern und Nutzern.

10.1 Erwartungsmanagement

„Erzähl mir einen Witz.“ Diese Aufforderung an einen Chatbot, der von einem Wetterdienst angeboten wird, dürfte zu Enttäuschungen bei Nutzern führen. Wie erfährt der Nutzer, was ein CUI oder ein intelligenter Assistent alles kann?

Bei klassischen Produkten ist das einfach. Man nimmt im Laden die Verpackung in die Hand und liest die Leistungsbeschreibung, oder man sucht vorab Ähnliches im Internet. Aber ein CUI kommt ohne Verpackung und Leistungsbeschreibung. Es muss intuitiv funktionieren, niemand liest das Handbuch. Es macht wohl auch wenig Sinn, den CUI das Handbuch vorlesen zu lassen.

Es ist eine Frage des Erwartungsmanagements. Der Anbieter eines CUIs muss dem Nutzer auf geeignetem Weg klar kommunizieren, dass er oder sie mit einer Maschine spricht. Dass es Limitierungen bei Funktionen und auch beim Verstehen von Sprache (bei Dialekten, Nuscheln, Slang etc.) gibt. Die Erwartung dürfte unter anderem so hoch sein, da bereits seit Jahren Diktate durch Automatic Speech Recognition in etlichen Bereichen gut funktionieren. Bei bestimmten Berufsgruppen wie Ärzten, Anwälten oder Steuerberatern klappt das aufgrund der stets wiederkehrenden, formelhaften Sprache extrem gut. Aber auch die Diktiersysteme in Smartphones liefern aus dem Stand, ohne dass es auf die Aussprache des Nutzers trainiert werden muss, bereits brauchbare Ergebnisse. Dies führt zu der Erwartung: Das System versteht mich, warum liefert es mir keine nützlichen Antworten? Das Erkennen des gesprochenen Wortes und seine Umsetzung in Schriftsprache

über Natural Language Processing (NLP) ist die eine Seite. Die Umsetzung der Aufforderung in eine brauchbare Antwort oder einen Service ist die andere – deutlich komplexere – Seite.

Eine Form des Erwartungsmanagements nennen die Fachleute „Fail Gently“. Bei Fragen oder Aufgaben, die ein CUI nicht versteht oder nicht beantworten kann, führt man den Nutzer in den Antworten behutsam zum eigentlichen Thema zurück. So schafft es der Wetter-Bot um den Witz herum und liefert eine Aussicht auf die Wetterlage der kommenden Tage.

10.2 Herausforderung Multimodalität

Noch fehlt CUIs weitestgehend die Multimodalität. Nur wenn der Anwender über sämtliche Eingabegeräte hinweg seine Wünsche äußern kann und am Ende zum Beispiel eine Produktauswahl auf einem Monitor angezeigt bekommt, werden CUIs Teil des Alltags. Um es etwas konkreter auszudrücken: Der Anwender spricht den Wunsch nach einem neuen Kleidungsstück in seinem Wohnzimmer. Zwar kann er oder sie Farbe, Schnitt und Größe genau definieren, aber man sieht die Auswahl nicht. Kann das CUI dem Nutzer eine Auswahl auf sein Tablet senden? Übernimmt das CUI im Auto die Auswahl, falls man während der nächsten Fahrt die Auswahl noch enger eingrenzen möchte? Wie übergibt man sprachbasierte Suchen von einem Gerät zum anderen? Wer wird diese Rolle ausfüllen? Der Gerätehersteller durch sein Betriebssystem oder der CUI-Anbieter? Bleibt es bei den Walled Gardens einzelner Anbieter? Gibt es ein Betriebssystem oder eine Plattform, auf der sämtliche sprachbasierte Kommunikation läuft? Fragen, die wir in Kap. 9 ausführlich behandelt haben, die bis heute aber nicht befriedigend gelöst sind und für eine Vielzahl unvollständig funktionierender und schlecht integrierter Dienste verantwortlich sind.

10.3 Woher stammt das notwendige Wissen?

Wissen als Grundlage für die CUIs ist ausreichend vorhanden. Das bezieht sich vor allem auf faktisches Wissen, also Informationen aus Quellen wie Wikipedia, oder Produkteigenschaften. Über Schnittstellen (APIs) können CUIs auf Datensammlungen von Unternehmen, Plattformen, Serviceangeboten und auch Suchmaschinen zugreifen.

Die Herausforderung liegt viel mehr in der Aufbereitung des Wissens. Bislang wird Wissen für die Darstellung auf Bildschirmen aufbereitet. In den Datenbanken finden sich Größen, Abmessungen, Farben sowie Materialangaben eines Produktes. Doch wenn ein Anwender sagt: „Ich möchte selber Eis machen.“, muss die Maschine aus dem Substantiv „Eis“ und dem Verb „machen“ den Intent herauslesen. Es geht hier nicht um Produkteigenschaften wie Watt oder Füllmengen, sondern um die Anwendung. Der Fragende benötigt also eine Eismaschine für seine Küche. Gleicher gilt für Reisewünsche. „Ich möchte nach Rom.“ Die Aussage ist nicht schwer zu deuten. Die Maschine weiß durch die

Anbindung an eine Geo-Datenbank, dass Rom die Hauptstadt von Italien ist. Mit Sicherheit ist durch das Nutzerkonto bzw. die GPS-Daten des Smartphones auch bekannt, wo sich der Fragende aufhält. Doch bei der Anfrage wurde kein Verkehrsmittel spezifiziert. Die Maschine muss eigenständig über Schnittstellen (API) zu Flug-, Bus- und Bahngesellschaften mögliche Reiseverbindungen und Preise abfragen.

Derartige Schnittstellen sind die technische Voraussetzung für sämtliche Produkt- und Serviceangebote. Oft haben Unternehmen über Jahre hinweg eine hermetische Software-Architektur aufgebaut. Die Daten liegen in unterschiedlichen und abgeschotteten Silos. Die Öffnung nach außen über Application Programming Interfaces (APIs) ist jedoch für eine Anbindung an ein CUI zwingend notwendig. Vermutlich ist dies nur möglich, wenn Unternehmen ihre Infrastruktur konsequent auf so genannte Webservices umbauen, was kosten- und zeitaufwendig ist. Nur wer seine Produkte und Services auf diese Art zugänglich macht, hat in einer Speech Economy eine Chance.

Wir bewegen uns von einer Faktensuche zu einer semantischen Suche. Wie unterschiedlich das Ergebnis aussieht, verdeutlicht die Suchmaschine Wolfram Alpha. Sie selbst versteht sich bereits als semantische Wissensmaschine. Es geht weniger um das Auffinden von Fakten, sondern um die Beantwortung von Fragen. Eine Kombination aus Natural Language Understanding, Curated Data und Dynamic Algorithmic Computation liefert brauchbare Antworten. Damit ist Wolfram Alpha dem, was ein CUI leisten muss, schon deutlich näher als sonstige Suchmaschinen. Die Eingabe „run 58 Minutes and 7 km“ liefert zunächst eine Nachfrage nach dem Geschlecht und dem Körpergewicht. Danach liefert Wolfram Alpha mehrere Tabellen, die Aufschluss darüber liefern, wie viele Kilokalorien und Fett bei diesem Lauf verbrannt werden. Auch der Sauerstoffverbrauch sowie eine Spannbreite der Herzschläge pro Minute (in Abhängigkeit vom Alter) präsentiert die Wissensmaschine für die Joggingrunde.

Ihren praktischen Nutzen beweisen CUIs beim Kauf von Produkten oder Dienstleistungen. Gute und umfassende Beratung ist eine wichtige Voraussetzung für Transaktionen. Kunden erwarten zurecht, dass in einem Beratungsdialog selbstverständlich alle denkbaren Informationen vorgetragen werden.

Beispieldialog

- CUI: „Bundesschatzbriefe werden vom Deutschen Staat ausgegeben und man kann diese direkt beim Staat kaufen in ein Online-Depot buchen.“
- Nutzer: „Wie hoch sind die Servicegebühren?“
- CUI: „Die zu berücksichtigenden Kosten sind die Servicegebühren selbst, auch die laufenden Depotgebühren. Außerdem müssen Käufer – abhängig von der individuellen Konstellation – auch Einkommens- und in manchen Fällen auch andere Steuer bedenken.“
- Nutzer: „Sind Bundesschatzbriefe wirklich sicher?“
- CUI: „Die Wertpapiere sind vom Deutschen Staat garantiert und zusätzlich gibt es Absicherungen im Falle, dass die depotführende Bank in Schwierigkeiten geraten sollte.“

Von einem solchen Dialogverlauf sind wir allerdings noch recht weit entfernt. Zwar gibt es unzählige Versuche von Banken in den USA und Europa derartige Beratungsgespräche zu automatisieren. In der Realität allerdings verlaufen solche Dialoge noch viel weniger „informiert“, sondern eher schematisch, also nach den für bestimmte Keywords hinterlegten Satzbausteinen.

10.4 Stockende Dialoge

Eingangs haben wir die Besonderheit von Google Duplex betont. Der CUI kommuniziert extrem menschlich, in dem er ein „hmmm“ oder „aha“ in den Satz einbaut. Zu menschlicher Kommunikation gehört aber noch viel mehr: Abschweifungen, Pausen, Schlussfolgerungen und Bezüge zu vorher gesagten Dingen.

Bevor wir dazu kommen, sollen aber noch die technischen Herausforderungen wie zum Beispiel das Herausfiltern von Nebengeräuschen erwähnt werden. In der Wohnung mag im Hintergrund der Fernseher laufen oder ein Kind schreien. Die Fortschritte im Bereich Far-Field-Sound-Erfassung [2] – die Technologie, die für den Umgang mit Störgeräuschen zuständig ist – ist verblüffend und bildet eine der Grundlagen dafür, dass Smartspeaker überhaupt im Alltag gebrauchsfähig geworden sind. Amazon scheint in diesem Feld einen interessanten Vorsprung erreicht zu haben. Im Straßenverkehr oder auf dem Bahnsteig ist ein bunter Chor an Geräuschen, die vom eigenen Sprachbefehl separiert werden müssen. Aber jeder Sprecher ist anders: Betonung, Dialekt, Slang und Abkürzungen stellen ein CUI vor Herausforderungen. Wann ist ein gesprochener Befehl beendet? Es gibt keine hörbare Interpunktions. Eventuell spricht der Nutzer nach zwei, drei Sekunden Pause weiter und konkretisiert seinen Wunsch.

In manchen Fällen wird ein und dasselbe Wort unterschiedlich ausgesprochen. Bitten Sie einmal einen Süddeutschen Chemie oder China auszusprechen. Wiederholen Sie diesen Versuch in Norddeutschland. Oder nehmen wir die Houston-Street in Manhattan. Man könnte meinen, sie wird wie der Name der Stadt in Texas [justen] ausgesprochen. Weit gefehlt, New Yorker sagen [hausten].

Eine weitere Herausforderung ist der Dialog-Flow, also der Bezug der Sätze zueinander. Gehören der vierte und fünfte Satz, die erst mit einigen Sekunden Unterbrechung ausgesprochen wurden, noch zum ursprünglichen Wunsch oder ist das bereits ein neuer Befehl? Wie geht ein CUI mit Anschlussfragen um? Gehört die nächste Frage noch zur vorangegangenen oder geht es um ein neues Thema?

Beispieldialog

- Nutzer: „Wie spät ist es in New York?“
- CUI: „13:30 Uhr“
- Nutzer: „Und in Los Angeles?“
- CUI: „Los Angeles ist mit 3.976.322 Einwohnern die größte Stadt im US-Bundesstaat Kalifornien. Sie liegt am Pazifischen Ozean und am Los Angeles River.“

Jede NLP-Lösung muss innerhalb einer Chat-Session den so genannten State erhalten. Das ist sozusagen das Gedächtnis eines CUIs. Je mehr Informationen im State landen, desto komplexer fällt die Analyse aus. Menschen haben gelernt, kommunikative Bezüge zu erkennen. Ein CUI muss genau wissen, wann ein Befehl abgeschlossen ist. Menschen wissen das u. a. durch non-verbale Signale wie Kopfnicken oder Handgesten.

Typische menschliche Kommunikationsmuster wie Pausen, Abschweifungen, Gedankensprünge, Schlussfolgerungen und Bezüge zu vorher geäußerten Wünschen werden in vielen Fällen noch nicht richtig erkannt.

Nachfragen rückgängig zu machen, ist noch unmöglich. Den letzten Eingabeschritt über die Tastenkombination Control + Z oder einen Zurück-Pfeil ist für Computernutzer das Normalste der Welt. Doch wie kommt man bei einem CUI zurück? Wie korrigiert man den zuletzt gesprochenen Befehl, wenn man einen Fehler gemacht hat?

Sprechpausen, in denen man kurz nachdenkt oder einen Bezug auf Dinge herstellt, die vor einigen Minuten gesagt wurden, sind für Maschinen ein Problem. Ein CUI muss wissen, wann ein gesprochener Befehl beendet ist und die Analyse beginnen kann. Um Bezüge herstellen zu können, muss die Maschine einen Dialog speichern. Je länger der Dialog, je größer der gespeicherte State, desto schwieriger ist die Auswertung. Wie umfangreich muss das „Kurzzeitgedächtnis“ einer Maschine sein?

Beispieldialog

- Anwender: „Ich möchte in den Sommerferien nach Spanien ans Meer in den Urlaub fahren.“
- CUI: „Ok, ich suche passende Angebote für Spanien. Zu welchem Zeitraum?“
- Anwender: „Mitte Juli bis Mitte August.“
- CUI: „Möchtest Du ab Deinem Wohnort fliegen?“
- Anwender: „Ja.“
- CUI: „Nach welcher Unterbringung soll ich suchen? Hotelzimmer, Appartement oder Ferienhaus?“
- Anwender: „Appartement für vier Personen.“
- CUI: „Soll Verpflegung mit dabei sein?“
- Anwender: (Pause) „Ach ne, ich glaube Griechenland gefällt mir dieses Jahr doch besser.“

Ein Intent ist die Absicht eines Nutzers. Die Kernaufgabe eines CUIs ist herauszufiltrieren, worin der Intent besteht. Enthält nun ein Satz mehrere Intents sind die meisten CUIs noch überfordert.

Beispiel

„Ist ein mexikanisches Restaurant in der Nähe, das jetzt geöffnet hat, dann reservier mir für 13 Uhr einen Tisch für vier Personen.“

Die meisten Dialoge in der realen Welt mit heute verfügbaren Sprachassistenten scheitern an den hier beschriebenen Herausforderungen. Natürlich arbeiten die Anbieter intensiv daran, damit ihre NLP- und Dialogsysteme auch mit Pausen, wechselnden Dialekten und stückweise geäußerten Multi-Intents zurechtkommen. Vollkommene Zukunftsmusik allerdings ist ein Feld, das im zwischenmenschlichen Dialog selbstverständlich und ausgesprochen wichtig ist: Menschen warten mit der Verarbeitung des vom Gegenüber Geäußerten ja nicht, bis dieser aufgehört hat zu sprechen. Im Gegenteil: Wir stimmen unserem Gesprächspartner durch Mimik, Gestik und auch durch bejahende Äußerungen zu, oder wir widersprechen eben. Conversational User Interfaces, die dies imitieren könnten, werden „schrithaltende Agenten“ genannt. Ihre Architektur ist naturgemäß komplex, erfordert hohe Verarbeitungsgeschwindigkeit, wie sie nur bei On-Device-Dialogsystemen bereitgestellt werden kann, und sind daher leider noch nirgends in der Praxis verfügbar, was zu unnatürlich stockenden Dialogen führt. Es ist wie bei einer Unterhaltung über Funk. Immer jeweils nur der eine oder der andere Gesprächsteilnehmer kann sprechen. „Roger and over“.

10.5 Fehlende Emotionen

Gefühle spielen in menschlichen Dialogen eine wichtige Rolle. Mein Gegenüber sieht, wie ich mit den Händen gestikuliere, wie ich meinen Körper einsetze. Er erkennt an Tempo, Lautstärke und Intonation wie meine Gefühlslage ist und reagiert entsprechend. Wie gut kann ein CUI den Menschen einschätzen?

Hört ein menschlicher Assistent den Satz „*Ich muss nach Berlin.*“, weiß er oder sie anhand der Emotionalität, dass der Sprecher es eilig hat. Vermutlich will er sofort nach Berlin. Da ergibt es keinen Sinn, nach Zugverbindungen für morgen zu schauen. Erinnert sich das CUI beim Satz „*Ich muss wieder nach Berlin*“ daran, dass dieser Wunsch bereits vergangene Woche geäußert wurde? Hat es sich die Reservierungspräferenzen für die Bahn (BahnCard 50, Abteil, Fenster, 2. Klasse) und das Ziel Ostbahnhof gemerkt?

Natürlich arbeiten die Forscher daran, dass auch CUIs Emotionen in der Sprache richtig erkennen. Doch stecken diese Arbeiten in den Anfängen [3].

Die Reaktion eines CUI auf Gefühle des Nutzers kann auch kritisch gesehen werden. Zum Beispiel wenn der Bot eines Maklers auf das Wort „Scheidung“ reagiert und dem Nutzer unaufgefordert Ein-Zimmer-Wohnungen im Umkreis vorschlägt [4].

Auch in Bezug auf Nudging kann ein emotionales Verhalten kritisch gesehen werden. Nudging beschreibt die Einflussnahme hin zu einer gewünschten Handlung. Wobei der Handelnde nicht das Gefühl hat, von außen beeinflusst worden zu sein. Die Handlung beruht einzig und allein auf seiner Entscheidung.

Derzeit erkennen CUIs wenig oder überhaupt keinen Kontext. Sie wissen nicht, wo der Nutzer gerade ist, oder können vorhandenes Wissen nicht einbeziehen. In welcher Gefühlslage befindet er sich? Was geschah direkt vor der Anfrage? Je besser ein CUI seinen

Nutzer, seinen Tagesablauf, seine Vorlieben und Gewohnheiten kennt, desto besser werden Wünsche umgesetzt.

10.6 Mangelnde Mehrsprachigkeit

Ein CUI versteht immer nur eine Sprache. Natürlich kann man die Sprache wechseln, doch die Limitierung bleibt bestehen. Es wird immer nur die aktuell gewählte Sprache erkannt. In Deutschland kommt es oft vor, wenn man nach Musiktiteln, Bands, Schauspielern oder Filmtiteln sucht, dass man ins Englisch wechseln muss oder die Wörter englisch ausspricht. Damit sind viele CUIs bei der Erkennung überfordert. In USA sprechen etliche Menschen in einer Mischung aus Spanisch und Englisch, in Indien in Mischungen ihrer Muttersprachen und Englisch: man spricht von Hinglish und nahezu alle derzeit verfügbaren Systeme sind mit dieser weltweit sehr verbreiteten Sprachenmischung vollkommen überfordert.

10.7 Grobe Werkzeuge

Ein Blick auf die teilweise noch ernüchternde Realität muss auch die Werkzeuge beinhalten. Wer heute ein CUI entwickelt, muss – bildlich gesprochen – mit Steinzeitwerkzeugen arbeiten. Selbst die Tool-Angebote von marktführenden Unternehmen bieten elementare Funktionen wie beispielsweise eine Zusammenarbeit in Arbeitsgruppen nicht an. Die Mehrzahl der Entwicklungstools bietet keine Versionierungen. Die Werkzeuge für Prototyping und Testing von CUIs steckt noch in den Kinderschuhen. Selbst als Anwendungsanfänger gerät man schnell an die Funktionslimits dieser Werkzeuge, sodass anspruchsvolle Entwicklungen häufig unter den unzureichenden Tools leiden. Die Qualität der Arbeitsmittel bestimmt letztendlich die Perfektion des endgültigen Produkts, weshalb viele CUIs heute im Praxistest nicht vollständig überzeugen können.

10.8 Wo stehen wir auf dem Hype Cycle?

Die Erwartungen an CUIs sind hoch. Insbesondere bei intelligenten Assistenten erwartet der Nutzer Antworten und Lösungen auf Fragen und Herausforderungen in allen Lebensbereichen. Dazu soll das CUI auf abstruse, unverschämte oder vulgäre Anfragen möglichst smart und humorvoll antworten. Doch die CUIs stehen in der breiten, öffentlichen Nutzung erst am Anfang. Auch sie durchlaufen den Hype Cycle, den das Beratungsunternehmen Gartner für sämtliche technische Innovationen entworfen hat (siehe Abb. 10.1). Fraglich ist nur, ob wir das „Tal der Enttäuschungen“ bereits in Richtung „Plateau der Produktivität“ verlassen haben oder noch auf dem „Gipfel der überzogenen Erwartungen“ stehen.

10.9 Voreingenommenheit und Fehler von Algorithmen

Amazon hat in den USA sein Gesichtserkennungsprogramm Rekognition an etliche Polizei- und Bundesbehörden verkauft. Die M.I.T. Media Labs haben das System eingehend untersucht und dabei kam heraus, dass es große Unterschiede zwischen Geschlechtern und Hautfarben gibt. Die Gesichter weißer, männlicher Personen wurden fast immer fehlerfrei erkannt und der richtigen Identität zugeordnet. Doch Frauen wurden in 19 Prozent der Fälle als Männer erkannt. Hatten die weiblichen Testpersonen einen dunklen Hautton, wurden sie zu 31 Prozent als Männer eingestuft [5].

Natürlich sind viele CUIs Bias und geben eine verzerrte Sicht auf die Welt wieder. Maschinelles Lernen benötigt Trainingsdaten, um Systeme weiter zu entwickeln. Hierbei werden Mainstream-Daten klar bevorzugt. Die Erklärung ist einfach: Die Maschinen sind Abbilder ihrer Schöpfer. Die meisten Entwickler sind jung, männlich und weißer Hautfarbe. Häufig kommen sie aus einem englischsprachigen Umfeld und sprechen die Hochversion der Sprache, also frei von Slang oder Dialekt. Somit stammen auch die meisten Trainingsdaten aus diesen Umfeldern.

Es dominieren zudem bei CUIs Sprachen, die von möglichst vielen Menschen gesprochen werden. Kleine Sprachgruppen sowie Dialekte werden marginalisiert. Ein Finne, der ein wenig nuschelt oder lispt und dazu noch einen regionalen Dialekt spricht, wird mit einem CUI zunächst enttäuschende Erfahrungen machen. Mozilla versucht das Problem zu umgehen, indem es die Basis seiner Trainingsdaten verbreitert. Jeder kann einige Sätze zur Erkennung beisteuern [6]. Auch in Ländern wie Indien gibt es Beispiele, wie INIDC-NLP [7], die gegen diesen Effekt ansteuern. In dem Land gibt es eine große Zahl eigenständiger Sprachen. Tamil ist eine davon, doch sprechen rund 200 Millionen Menschen

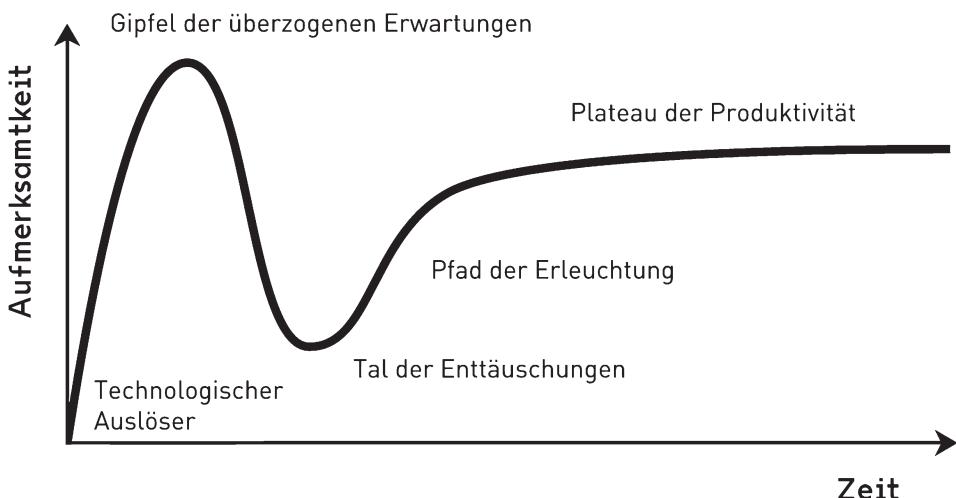


Abb. 10.1 Der Gartner Hype Cycle

diese Sprache. Von klein kann also keine Rede sein. Vermutlich kann man diese Herausforderung nur meistern, wenn man die Arbeit auf mehr Schultern verteilt. Crowdsourcing lautet die Antwort. So lässt das Start-up Hive [8] digitale Daten, in diesem Fall visuelle Daten für Verkehrserkennung, von der Crowd erkennen und labeln.

10.10 Datenschutz und Datensicherheit

Im Laufe der Zeit sammeln CUIs eine große Zahl gesprochener Befehle. Werden diese Aufnahmen von den Betreibern gespeichert? Sind sie gegen Zugriffe Dritter geschützt? Werden Vorlieben, Wünsche, Kaufgewohnheiten mit Nutzerzustimmung für ein personalisiertes Marketing ausgewertet?

Auf diese Fragen müssen Betreiber den Nutzern Antworten liefern. Microsoft hat beispielsweise sechs Prinzipien [9] entwickelt, die bei ihren CUIs gelten sollen:

1. Kontrolle: Der Anwender hat die Kontrolle über seine Daten. Einfach zu benutzende Werkzeuge und eine klare Auswahl erleichtern die Einstellungen zur Privatsphäre.
2. Transparenz: Microsoft informiert, was gesammelt wird. So kann der Anwender entscheiden, ob er das Produkt weiter nutzen möchte.
3. Sicherheit: Microsoft will Daten vor den Augen Dritter sichern und verschlüsseln.
4. Gesetzliche Vorgaben zum Datenschutz werden in jedem Land beachtet
5. Es wird kein Inhaltstargeting betrieben und es werden keine persönlichen Daten aus Chats, Mails oder Dateien genutzt, um zielgerichtet Anzeigen auszuliefern.

Das hört sich alles sehr gut an. Beim 6. Punkt allerdings zeigt sich bereits, wie schwierig derartige Selbstverpflichtungen in der Wirklichkeit umzusetzen sind. Im letzten Punkt spricht Microsoft davon, dass keine Daten gesammelt werden. Falls doch, dann nur um die Anwendererfahrung zu verbessern.

In der Öffentlichkeit wurden alle Anbieter von CUIs heftig dafür kritisiert, dass Gesprächsprotokolle, bei denen das CUI nicht zur Zufriedenheit des Nutzers funktioniert hat, zur Auswertung und Verbesserung von Menschen abgehört und überprüft wurden. Für viele Experten eine schwer verständliche öffentliche Aufregung, denn wie sollen derartige Systeme verbessert werden, wenn nicht durch gründliche Analyse der Fehler? Dennoch ist es für viele Nutzer ein weiterer Beleg dafür, dass sprachbasierte Systeme nicht sicher sind. Eine Sorge, die eine Mehrzahl von Nutzern ohnehin schon beschäftigt. In einer ebenfalls von Microsoft 2019 durchgeführten Studie [10] sagen über 41 Prozent der Befragten, dass sie sich unwohl fühlen, da Smartspeaker sie aktiv belauschen. Damit würden Informationen gesammelt, die nicht privat gehalten werden und möglicherweise für andere Zwecke genutzt werden. Über 50 Prozent der Studienteilnehmer sagten ihre Daten seien nicht sicher.

Wie bei jeder technischen Neuerung kommen mit CUIs und intelligenten Assistenten neue Sicherheits-Fragestellungen, die man ernst nehmen muss. Dies vor allem auch vor dem Hintergrund, dass CUIs sich genauso nahe an den Nutzern befinden, wie Smartpho-

nes, die ja in der Tasche am Körper überall mitgetragen werden und am Abend nicht selten auf dem Nachttisch neben dem Bett liegen. Wenn CUIs und intelligente Assistenten funktionieren sollen, dann müssen diese mit sensiblen persönlichen Daten agieren und sich mit möglichst vielen Diensten (Kalender, Kontakte, Mail usw.) verbunden sein. Auch das Smartphone trägt viele dieser Daten mit sich.

Ein zusätzliches Risiko liegt bei im Smartphone arbeitenden CUIs in der Möglichkeit der ungewollten Stimmaktivierung. Generell sind die Stimmen von Menschen verschieden, wie ein Fingerabdruck. Daher gibt es schon seit längerem biometrische Verfahren, die Stimmen am Telefon erkennen können. Amerikanische Banken wie Charles Schwab Voice ID [11] arbeiten bereits mit dieser Technologie zur Verifizierung von telefonbasiertem Online-Banking. Allerdings ist die Integration dieser Technologien in den derzeit verbreiteten Diensten noch nicht weit fortgeschritten [12]. Werden diese allein als Identifikation eingesetzt, können sie unter Umständen durch einfache Stimmaufnahmen oder komplexe Sprachsyntheseverfahren übergegangen werden. Betrachtet man das Gesamtbild der Risiken, entstehen Risiken im Netzwerk der Plattformen selbst. An jeder Stelle des Netzwerkes können durch Kriminelle oder staatliche Stellen Datenabhörstationen errichtet werden. Zwar ist bislang noch von keinen massiven Manipulationen auf dieser Ebene [13] berichtet worden, was aber nicht unbedingt bedeuten muss, dass dies nicht bereits stattgefunden hat und von den Betreibern nicht publiziert wurde. Oder eventuell zukünftig stattfinden wird.

Durch Analyse des Datenverkehrs am heimischen Router, kann entgegen der Sorge der Öffentlichkeit selbst ein Laie sehen, dass ein Smartspeaker tatsächlich erst nach Aktivierung durch den Nutzer die Mikrofone einschaltet und Sounddateien in die Cloud zur Analyse und Verarbeitung schickt. Natürlich können Smartspeaker von Kriminellen oder staatlichen Stellen an der Hardware manipuliert und damit zu Abhörstationen werden. Eine Sorge, die man allerdings bei allen technischen Geräten mit eingebauten Mikrofonen (Telefonen, Smartphones) haben muss (Abb. 10.2).

10.11 Adversarial Attacks

Ein von der Öffentlichkeit noch wenig wahrgenommenes Risiko besteht darin, dass Sprachsysteme getäuscht werden und für Menschen nicht wahrnehmbare Signale empfangen. Dies sind so genannte Adversarial Attacks (siehe Abb. 10.3). Die ersten Beispiele stammen aus dem Bereich der Bilderkennung. Optische Systeme werden trainiert, in dem die Entwickler vorgeben, wie beispielsweise ein Panda aussieht: Tier, Fell, schwarz-weiß, große Augen. Danach erhält das System zu Trainingszwecken möglichst viele weitere Bilder von Pandas und lernt so. Wobei lernen hier natürlich anders funktioniert als bei Menschen. Das System ermittelt eine Wahrscheinlichkeit, nach der die Bildpunkte einen Panda darstellen. Ab einer bestimmten Schwelle, die der Entwickler festlegt, entscheidet die Bilderkennung: Dies ist ein Panda.

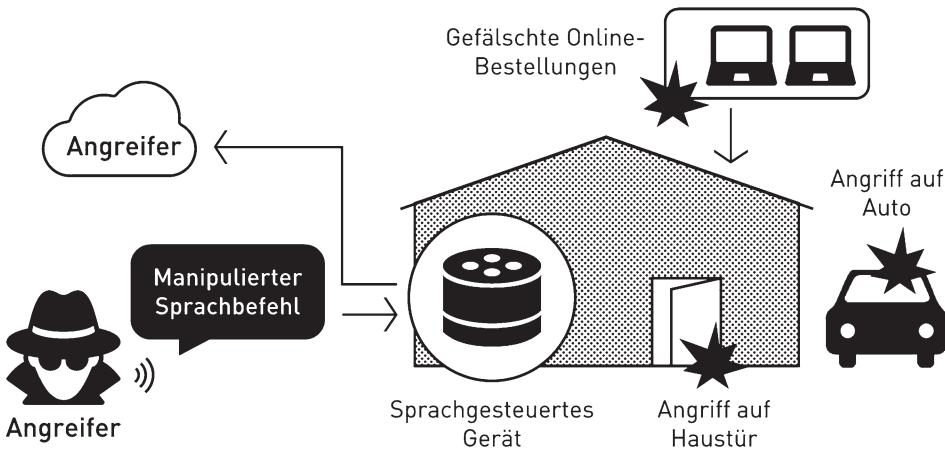


Abb. 10.2 Adversarial Attacks bei CUIs

Nun ist es nicht sonderlich schwer, die Bilderkennungssysteme zu täuschen. Indem man zusätzliche Objekte oder auch nicht sichtbares Bildrauschen bzw. Farbmuster in das Bild einfügt. So kann man die Bilderkennung täuschen. Plötzlich ist für das System ein Gibbon statt ein Panda im Bild zu erkennen.

Die Spracherkennung eines CUIs ist vereinfacht gesagt wie bei Bildern eine Mustererkennung mit Wahrscheinlichkeiten. Die Angreifer verändern hierbei das Audiomuster, da vom System erkannt wird. Das erfolgt über Sounddateien, die sich für das menschliche Gehör wie Rauschen oder Störgeräusche anhören (siehe Abb. 10.3). Läuft ein derartiger Sound im Hintergrund einer Radio- oder TV-Wiedergabe während ein Sprachbefehl erteilt wird, erkennt das CUI einen anderen Befehl als vom Anwender gesprochen. Damit lässt sich eine ungewollte Bestellung bei einem Versandhandel auslösen. Es werden Schlagzeilen aus Zeitungen oder Radioberichte manipuliert. Oder Angreifer deaktivieren auf diesem Weg die Sicherheitskameras und entriegeln die Haustür. Um die Funktionsweise der Adversarial Attacks im Bereich der CUI zu verstehen, empfehle ich die Hörbeispiele auf der Seite adversarial-attacks.net.

CUIs müssen gegen derartige Angriffe gesichert werden. Experten sind sich einige, dass die Lösung in der Nutzung besserer Trainingsdaten liegt, was angesichts der Fülle denkbarer Angriffspunkte eine große Herausforderung ist. Noch gibt es keine Universal-Methode, um Adversarial Attacks zu unterbinden. Doch mit Blick auf das weite Feld der Malware mit Phishing-Attacken, Trojanern und Viren liegt die Vermutung nah, dass es auch hier auf ein ewiges Hase-und-Igel-Rennen hinauslaufen wird.

Je mehr wir von Assistenten bedient werden wollen, je mehr Komfort wir durch eigenständige Entscheidung der Agenten genießen wollen, umso mehr werden wir offensichtlich verletzlich. Die Antworten auf die vielfältigen Fragen von Datenschutz und Datensicherheit sind längst nicht alle geklärt und gelöst. Blickt man auf Adversarial Attacks, kommen im Gegenteil täglich weitere Herausforderungen auf die Anbieter zu.

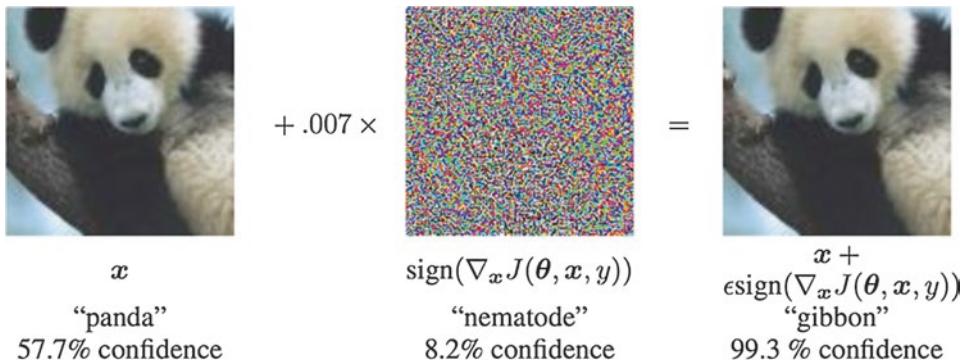


Abb. 10.3 Optische Störmuster beeinflussen die Bilderkennung (Foto: Goodfellow et al. ICLR 2015)

10.12 Schnelle Schritte

Nach dem Lesen könnte der Reality Check den Eindruck erwecken, CUIs sind noch extrem fehlerbehaftet und unfertig. Doch dies ist nur eine Momentaufnahme. Wir werden enorme Entwicklungsschritte erleben. Viele der hier genannten Einschränkungen werden in kürzester Zeit überwunden sein. Das New Yorker Risikokapital-Unternehmen Loup-ventures untersucht im Jahresrhythmus die Verbesserungen bei den großen intelligenten Assistenten [14] (siehe Abb. 10.4). Dazu werden jedem der drei Anbieter, Google Assistant, Alexa und Siri, 800 Fragen in fünf Kategorien gestellt. Die Grafik zeigt, zu welchem Prozentsatz korrekte Antworten geliefert wurden. Interessant ist ein Blick auf die Veränderung im Vergleich zum Test im Jahr 2018. Danach hat Alexa 18 Prozentpunkte aufgeholt. Siri verbesserte sich um fünf und Google Assistant um sieben Prozentpunkte bei den richtigen Antworten.

Wem der Glaube an die Entwicklung bei CUIs fehlt, erinnere sich bitte an folgende Entwicklungen: Bis sich der Online-Ableger des Nachrichtenmagazins „Der Spiegel“ zum führenden Nachrichtenportal im Web etabliert hatte, vergingen rund zehn Jahre. In den Anfangstagen des App-Stores und des iPhones war die Bier-Trinker-App Gesprächsthema auf jedem Büroflur. So zu tun, als trinke man ein Bier aus einem mehrere hundert Euro teuren Smartphone, ist nüchtern betrachtet albern. Doch es war eine spielerische Annäherung an die Möglichkeiten eines mobilen Taschencomputers. Dieses Gerät war erstmals mit einem berührungsempfindlichen Bildschirm ausgestattet und hatte mit dem Gyroskop einen Sensor eingebaut, den man wahlweise als Wasserwaage oder zur Anzeige von Flüssigkeitspegel nutzen konnte. Innerhalb von rund fünf Jahren hat sich das Smartphone zum Universalgerät entwickelt. Auf Reisen (Fahrkarten, Bordkarten, Hotelreservierung), zu Hause (Smart-Home-Steuerung), zur Medienwiedergabe (Musik und TV-Streaming), beim Sport (Tracking) und als meistgenutzte Fotokamera leistet es heute unverzichtbare Dienste.

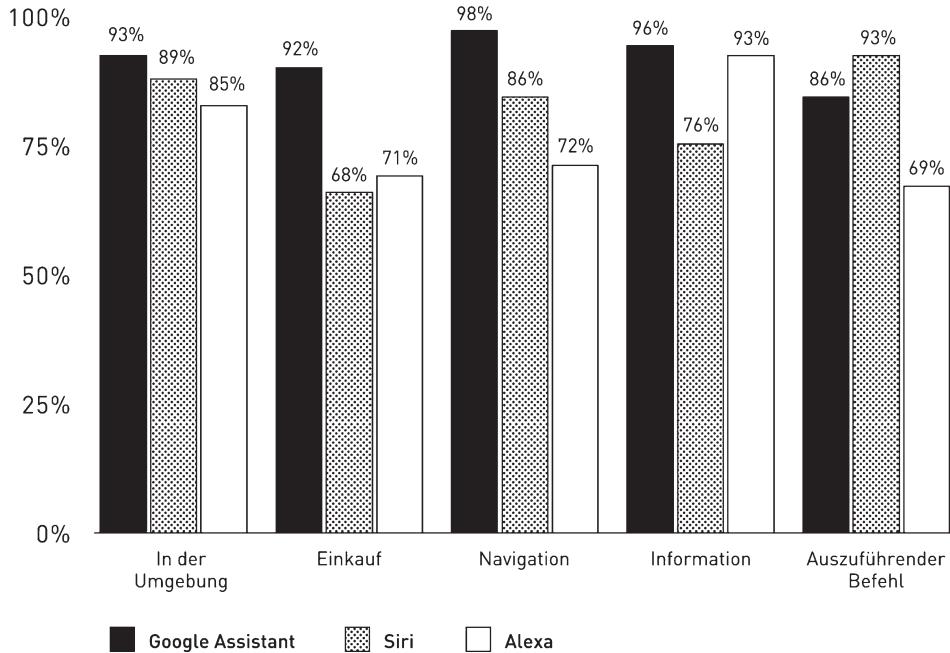


Abb. 10.4 Prozentsatz korrekter Antworten auf Fragen in unterschiedlichen Kategorien

Im Bereich der CUIs wird es noch schneller gehen. Die Investitionen der großen Player Google und Amazon sind dafür ein untrügerisches Indiz. Ein Blick in die Stellenausschreibungen von Amazon zeigt, dass für den Bereich Alexa weltweit über 2100 Positionen unbesetzt sind. Werden diese Arbeitsplätze besetzt, dürfte derart geballtes Entwicklungs-Know-how innerhalb kürzester Zeit zu beeindruckenden Ergebnissen führen.

Literatur

1. Perez, Sarah: Google launches CallJoy, a virtual customer service phone agent for small business, (01.05.2019), <https://techcrunch.com/2019/05/01/google-launches-calljoy-a-virtual-customer-service-phone-agent-for-small-businesses/>, zugegriffen 12.11.2019
2. Tilley, Aaron: Amazon Opens Up Echo's Far-Field Voice Recognition Tech To Outsiders, (13.04.2017), <https://www.forbes.com/sites/aarontilley/2017/04/13/amazon-opens-up-echos-far-field-voice-recognition-tech-to-outsiders/#6d572d883c2d>, zugegriffen 04.11.2019
3. Paulus, Wolf: Conversational UI: It's Not Just What You Say, But How You Say It, (09.04.2019), <https://www.intuit.com/blog/technology/conversational-ui-its-not-just-what-you-say-but-how-you-say-it/>, zugegriffen 03.11.2019
4. Levy, Dan: The power and potential of emotional chatbots, (19.10.2018), <https://www.marketing-techniques.net/news/2018/oct/19/power-and-potential-emotional-chatbots/>, zugegriffen 20.11.2019

5. New York Times, Amazon Is Pushing Facial Technology That a Study Says Could Be Biased, Natasha Singer, 61.2019
6. Mozilla Common Voice Project, <https://voice.mozilla.org/de>, zugegriffen 12.11.2019
7. IndicNLP, AI for Indian languages, <https://indicnlp.github.io/>
8. Wiggers, Kyle: Hive taps a workforce of 700.000 people to label data and train AI models, (16.11.2018), <https://venturebeat.com/2018/11/16/hive-taps-a-workforce-of-700000-people-to-label-data-and-train-ai-models/>, zugegriffen 16.11.2019
9. 2019 Voice Report: Consumer adoption of voice technology and digital assistants, (April 2019), Seite 24, <https://about.ads.microsoft.com/en-us/insights/2019-voice-report>, zugegriffen 23.09.2019
10. 2019 Voice Report: Consumer adoption of voice technology and digital assistants, (April 2019), Seite 24, <https://about.ads.microsoft.com/en-us/insights/2019-voice-report>, zugegriffen 23.09.2019
11. Schwab voice ID Service, https://www.schwab.com/public/schwab/nn/m/voice_biometric.html, zugegriffen 20.11.2019
12. Floemer, Andreas: Alexa kann jetzt auch Stimmenkennung – „Mein Stimme“ startet in Deutschland, (08.12.2018), https://t3n.de/news/alexa-meine-stimme-stimmenkennung-deutschland-1131090/?_, zugegriffen 23.11.2019
13. Pursche, Olaf: Amazon Echo Dot und Alexa App im Test, (24.2.2017), <https://www.iot-tests.org/de/2017/02/amazon-echo-dot-und-alexa-app-im-test/>, zugegriffen 23.11.2019
14. Thomson, Will; Munster, Gene: Annual Digital Assistant IQ Test, (15.08.2019), <https://loupventures.com/annual-digital-assistant-iq-test/>, zugegriffen 23.11.2019



Zusammenfassung

Sprache ist die natürlichste Form menschlicher Kommunikation. Doch wollen wir immer und überall mit Maschinen reden? Im öffentlichen Raum kann es schnell unangenehm werden, wenn andere mithören. Ein Stück weiter gedacht: Muss der Mensch für einfache Dinge überhaupt noch mit einer Maschine reden? Können nicht Maschinen mit Maschinen reden, wenn es um Hausputz, das Einkaufen oder die Verabredung für einen Termin geht? Aus Wearables werden Awareables: Die Sensorik in Uhren, Armbändern, Halsketten und Smartphones erfassen so viele Daten von und über uns, das vorausschauendes Handeln möglich wird. Da muss gar kein Wort mehr gesprochen werden.

11.1 Sprache als universelles Interface

Wie bereits am Anfang des Buches dargelegt, können natürlichsprachige Interfaces dazu dienen, intelligente Assistenten zu bedienen. CUIs können aber auch dazu eingesetzt werden, gebräuchliche grafische Benutzeroberflächen von Computern, Tablets und Smartphones mittels gesprochener Sprache zu steuern. Apples Siri hat dies bei Markteinführung im Jahr 2011 gezeigt.

Insbesondere mit den Möglichkeiten von Spracherkennung, die direkt auf dem Gerät erfolgt (On-Device-NLP), ergeben sich dabei vielversprechende neue Einsatzgebiete. Sobald Spracherkennung lokal und damit schnell erfolgt, ergibt es plötzlich Sinn, statt mit den Fingern auf dem Smartphone nach der richtigen App und dann der richtigen Funktion zu suchen, den entsprechenden Befehl einfach per Sprache einzugeben und das Gerät diesen ausführen zu lassen. Google zeigt mit seiner auf der aktuellen Smartphone-Generation Pixel installierten Android-Version, was möglich ist – die Sprachbedienung

des gewöhnlichen Touch-Interfaces ist wirklich sehr beeindruckend. Wenig überraschend ist aber auch Apple in diesem Feld sehr aktiv.

Dass Sprachsteuerung keineswegs nur Befriedigung von Bequemlichkeit dienen kann, sondern auch interessante neue Perspektiven für Menschen mit entsprechenden Mobilitäts einschränkungen bieten, zeigt Apple in dem Video „Voice Control für MAC und IOS“ [1]. Darin steuert ein stark bewegungseingeschränkter junger Mann seinen Mac virtuos per Sprache. Er öffnet dabei ein Bildbearbeitungsprogramm und einzelne Bilder, die er sortiert und bewertet. Das Apple Programm „Voice Control“ bietet dabei im Fall von mehreren Auswahlmöglichkeiten jeweils eine kleine Zahl als Nummerierung der Auswahlmöglichkeiten an, die vom Nutzer per Voice schnell angewählt werden kann. Durch die Geschwindigkeit, in der die Befehle on device erkannt und verarbeitet werden, macht die Bedienung per Voice offensichtlich Freude und ist damit weit mehr als ein Notbehelf. Bei einer On-Device-Implementierung geht es darum die Spracherkennungsalgorithmen und Spracherkennungsreferenzen so kompakt zu organisieren, dass diese wenig Speicherplatz benötigen und auch auf weniger leistungsfähigen Computerchips verarbeitet werden können. Es ist abzusehen, das On-Device-NLP in den kommenden Jahren einfach auf einem Computerchip fest installiert – embedded – sein wird und einfach und kostengünstig genug, dass ein solcher Computerchip in jedem Lichtschalter eingebaut werden kann.

11.2 Eignet sich jeder Anwendungsfall für Sprache?

Eine Frage, die einerseits technisch beantwortet werden kann, andererseits im Laufe der weiteren Entwicklung sicherlich immer wieder neu beantwortet werden muss, da sich erfahrungsgemäß mit neuen Interfaces auch die Anwendungsfälle entwickeln und die Gesellschaft neue Regeln der Akzeptanz entwickelt.

Technisch gesehen, stellen Umgebungs- und Hintergrundgeräusche in vielen Fällen eine ernste Hürde für Sprache dar. Insbesondere aber dann, wenn das Eingabegerät nicht, wie im Fall von Smartphones oder Kopfhörern mit eingebauten Mikrofon, nahe am Sprechenden ist, sondern, wie im Fall von Smartspeakern, oder anderen Eingabegeräten wie Kühlschränken, Lichtschaltern usw. in einiger Entfernung zum Sprechenden stehen.

Zwar verlief die Entwicklung im Bereich von so genannten Fernfeld-Mikrofonen und Fernfeld-Technologien in den letzten Jahren überraschend schnell, was die Gerätekategorie von Smartspeakern überhaupt erst möglich gemacht hat, dennoch gibt es naturgemäß geräusch-physikalische Probleme, die auch mit noch so guten Algorithmen, die gewollte Stimmen und ungewollte Geräusche trennen, nicht zu überwinden sein werden. Neben einer lauten Maschine, im Verkehr, oder in großen Menschenansammlungen sind Grenzen gesetzt, die Voice nicht brauchbar erscheinen lassen.

Auch das Problem der Vertraulichkeit von Gesagtem und Gehörtem, sobald andere Personen in der Nähe sind, lässt sich technisch nur unter Opferung von Beweglichkeit in Form von Kopfhörern und speziellen in Fingerringen und Armreifen eingebauten Mikrofonen, die Wispern erkennen, überwinden. Microsoft [2] und andere Unternehmen arbeiten

bereits an derartigen Technologien (siehe Abb. 11.1). Und Amazon hat im Herbst 2019 mit Alexa Loop einen Ring zur Spracheingabe vorgestellt.

Eine weitere Herausforderung sind domainspezifische Vokabeln. So ist die Fachabkürzung „KAD“ (Key Account Director) ausgesprochen nicht vom Wort „cat“ zu unterscheiden, so dass Consumer CUIs wie Siri oder Alexa sich immer für die Katze entscheiden würden. Oracle und andere Unternehmen entwickeln aus diesem Grund eigene AI-Sprachsysteme, die im spezifischen Unternehmenskontext trainiert werden können [3].

Doch selbst, wenn diese Entwicklungen erfolgreich sein sollten, bleibt dies eine Frage gesellschaftlicher Akzeptanz. Wie stark sich diese positiv entwickelt, kann man jedoch gut daran erkennen, wie selbstverständlich heute Menschen in gefühlten untauglichen Umgebungen mit anderen telefonieren und Zugabteile ungewollt ganz selbstverständlich am persönlichsten Geschehen beteiligen. Ein Verhalten, das noch vor 20 Jahren – also vor der allgemeinen Verfügbarkeit von Mobiltelefonen – vollkommen undenkbar gewesen wäre. Wir können also gespannt sein, wie diese Frage der Akzeptanz in 20 Jahren aussehen wird.

Gute Sprachassistenten sind imstande, komplexe Anfragen zu bearbeiten, aber selbst die besten Sprachassistenten sind schlecht darin, komplexe Antworten zu geben, in denen eine Vielzahl von Informationen oder Auswahlmöglichkeiten enthalten sind. Wenn man dies berücksichtigt, wird deutlich, dass sich die besten Voice-Anwendungen vermutlich unterwegs aufbauen lassen. Also immer dann, wenn man Hands-free, oder gar

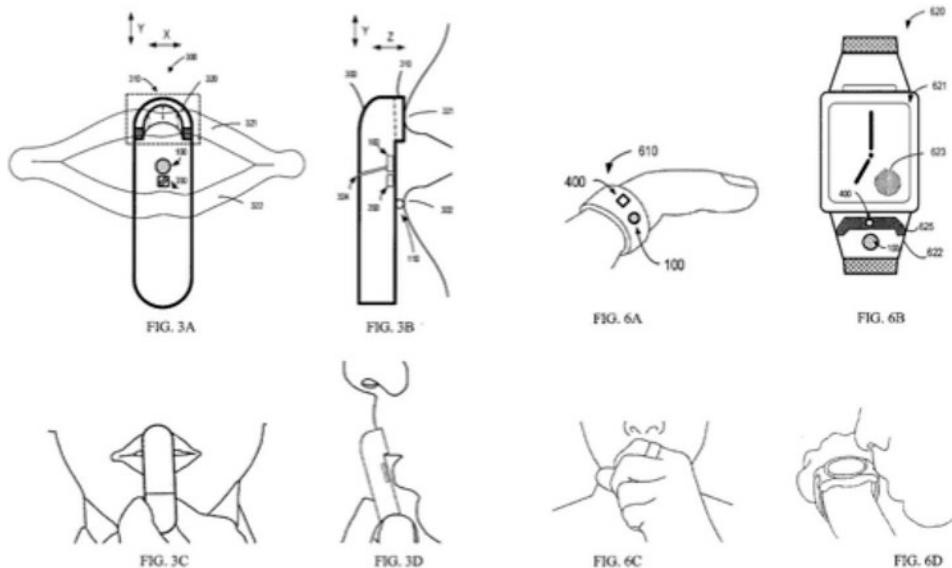


Abb. 11.1 Die Patentzeichnung beschreibt Spracheingabemethoden, die leises und für Umstehende nicht wahrnehmbares Flüstern und Tuscheln ermöglichen. (Foto: Microsoft)

Eyes-Free im Auto, auf dem Fahrrad, zu Fuß unterwegs ist und die Hände nicht frei hat und auch den Blick nicht vom Weg abwenden möchte. Mithilfe von Kopfhörern mit eingebauten Mikrofonen, wie sie Apple bereits seit längerem und nun auch Amazon und Google anbieten, lassen sich Fragen und Befehle relativ gut einsprechen und für eine Wegbeschreibung sind meist keine großen Auswahlmöglichkeiten notwendig. Dies scheint zwar im Widerspruch zum Umstand stehen, dass CUIs derzeit ja vor allem in Form von stationären Smartspeakern ihre größte Verbreitung finden. Wenn man aber beobachtet, mit welcher Intensität Amazon versucht, in den Markt der sog. „Hearables“ vorzustoßen kann man erkennen, dass diese Einschätzung offensichtlich auch beim derzeitigen Marktführer der Smartspeaker geteilt wird. Unter Hearables versteht man intelligente Kopfhörer, die Umgebungsgeräusche selektiv ausblenden können und für Spracheingabe optimierte Mikrofone eingebaut haben. Man kann feststellen, dass sich viele Anwendungsfälle nur dann für Voice eignen, wenn die Software Nebengeräusche filtern kann und die Anzahl notwendiger Schritte und Auswahlmöglichkeiten durch Software eingeschränkt werden kann. Je mehr also Steuerung durch natürliche Sprache stattfindet, desto mehr sind auch intelligente Assistenten, Softwareagenten oder Bots notwendig, um dies zu realisieren.

11.3 Eignet sich jeder Anwendungsfall für einen intelligenten Assistenten?

Die Antwort auf diese Frage ist im Grunde sehr einfach: Ja! Intelligente Assistenten können vermutlich in fast jedem denkbaren Anwendungsfall zu einem sinnvollen Einsatz gebracht werden. Strukturelle Hürden stellen hier die Bewältigung von Komplexität dar, denn um eine vermeintlich einfache Aufgabe teilautonom im Sinne des Nutzers abarbeiten zu können, bedarf es in der Regel sehr komplexer Prozesse und, wie schon mehrfach dargelegt, auch ausreichend Anbindung an Datenquellen aller Art.

Auch hier stellt sich neben der Frage nach der technischen Dimension auch die Frage nach der Akzeptanz. Es wurde in diesem Buch bereits mehrfach auf die zentrale Thematik Vertrauen eingegangen, das Nutzer intelligenten Assistenten entgegenbringen müssen. Diese Bereitschaft ist sicherlich in verschiedenen Altersschichten, aber auch Weltregionen, unterschiedlich ausgeprägt, weshalb auch die Eignung von Einsatzfeldern in verschiedenen Demografien und Märkten unterschiedlich zu bewerten ist.

Betrachtet man die Entwicklung von Anwendungsfällen bei Smartphones, so galten zu Beginn auch hier viele aus der PC-Welt bekannte Nutzungsszenarien zunächst nicht als smartphonegeeignet. Doch sie durchliefen über die Zeit verschiedene Metamorphosen bis es zu Smartphone-Banking oder Smartphone-Dating kam, die plötzlich auf dem Smartphone Sinn machten. Bekanntermaßen haben nicht alle Anbieter diese Metamorphosen erfolgreich durchlaufen. Das lässt Raum für neue, spezialisierte Anbieter und es entwickelten sich Anwendungsfelder, die nur mit dem Smartphone Sinn machen, wie z. B. Ride-Hailing und Car-Sharing-Angebote von Share Now, Moia und Uber.

11.4 Strukturelle Sweetspots

Obwohl es vermutlich keine Einschränkungen gibt, welche Anwendungsgebiete grundsätzlich für CUIs und intelligente Assistenten geeignet sind, gibt es doch Themenfelder, die besonders geeignet sind, die so genannten Sweetspots. Da Assistenten ihre Intelligenz durch Auswertung von historischen und aktuellen Daten lernen, sind Sweetspots alle Fragestellungen, in denen es große Bestände an Trainings und Referenzdaten gibt und hohe Fallzahlen es ermöglichen, schnelle Verbesserungszyklus zu durchlaufen. Im Idealfall kommen hohe Fallzahlen und eine geringe thematische Breite zusammen – teilweise bedingen sich diese ja auch.

Aus Nutzersicht sind für eine Erledigung durch intelligente Assistenten alle diejenigen Aufgaben attraktiv, die durch starke Hierarchie der Zwischenfragestellungen komplex und unangenehm erscheinen. Assistenten können im Idealfall ein Problem abschließend lösen, ohne neue, zusätzliche Fragestellungen aufzuwerfen. Bots finden, statt zu suchen. Softwareagenten können aus Nutzersicht sehr gut immer wiederkehrende Aufgaben erledigen.

Der Bereich, in dem all diese Bedingungen derzeit am besten zusammentreffen, sind Callcenter von Unternehmen, die im Kundenservice hohen Aufwand betreiben müssen: Telefongesellschaften, Versorger aller Art, Unternehmen mit Abo-Beziehung zu Kunden. In diesen Fällen, wie im After-Sales-Kundenkontakt bei Online-Händlern, gibt es historische Daten aus E-Mails und teilweise auch Chatprotokolle, die als Grundlage zur Auswertung und zum Training von intelligenten Assistenten genutzt werden können. Gleichzeitig ist die Vielfalt der Kundenbedürfnisse meist recht beschränkt: Vertrag kündigen, verlängern, erweitern, einschränken, technische Probleme und Probleme mit gelieferten Waren. 80 Prozent der Kundenanfragen betreffen 20 Kundenwünsche, und bei großen Anbietern gibt es sogar eine tägliche Fallzahl, die das System schnell besser machen kann. Daher sind Callcenter international derzeit einer der Sweetspots beim Einsatz von intelligenten Assistenten.

Im Folgenden soll es darum gehen, Einsatzgebiete von CUIs und intelligenten Assistenten vorzustellen, die zunächst vielleicht etwas abseits von diesen naheliegenden Anwendungen liegen. Es sollen inspirierende Beispiele aufgezeigt werden, bei denen Bots die Nutzer an etwas erinnern oder auf etwas aufmerksam machen. Es sollen Assistenten vorgestellt werden, die etwas für Menschen erledigen, oder diese überprüfen und dabei Sachverhalte erfassen. Es wird gezeigt werden, dass Bots bereits genutzt werden um Menschen mit Menschen zu verbinden und Nutzer aktiv in Dialoge zu verwickeln.

11.5 Software-Agenten, die erinnern und aufmerksam machen

Bots, die automatisierte E-Mails verschicken um auf bestimmte Ereignisse oder Angebote aufmerksam zu machen, sind schon seit längerem bekannt, ein Prinzip das auf Messenger-Dienste, wie WhatsApp zunächst einfach übertragen wurde und auch im Sinne von App-Notifications auftaucht. Ein Grundproblem besteht in der zunächst notwendigen Grundkonfiguration. Die Frage, worauf das System zu welchem Zeitpunkt aufmerksam

machen soll, muss meist in umständlichen Auswahlmenüs festgelegt werden und ist im Grunde bereits überholt, wenn die Auswahl bestätigt worden ist.

Conversational User Interfaces und intelligente Assistenten werden beim Bot-System von Capital One auf vielfältige Weise eingesetzt. Der Nutzer kann Eno [4] einfache Fragen stellen: „*Eno, wie viel habe ich in diesem Monat bei Starbucks ausgegeben?*“ und erhält die Antwort. „*Eno, benachrichtige mich, sobald Betrag X überschritten ist.*“ Auf diese Weise entsteht ein System von Benachrichtigungen, ohne dass der Nutzer dafür viel tun muss. Der Assistent fragt auch aktiv nach: „*Hast Du tatsächlich in diesem Monat zweimal bei H&M eine Bluse für 39,90 Euro gekauft?*“ und überprüft auf diese Weise im Zusammenspiel mit dem Kunden die Richtigkeit der Daten. Capital One ist Vorreiter beim Aufbau von Conversational User Services und der intelligente Assistent Eno ist auf verschiedenen Plattformen verfügbar, darunter Amazon Alexa und Microsoft Cortana.

Derartige Anwendungsfelder werden nicht nur im Umfeld von Privatnutzern entwickelt. Auch SAP und weitere Business-Intelligence-Softwareanbieter entwickeln Systeme, die es einfach machen, Benachrichtigungen für Ereignisse, Zusammenstellungen immer wiederkehrender Reports und andere im Geschäftsleben entscheidende Themen in der unendlichen Datenflut zu finden und für Entscheidungsprozesse einsetzbar zu machen. Wie viel einfacher ist es zu fragen: „*Hello Datenbank, sag mir die Mobile-Nutzer am Montag*“ und die Antwort zu erhalten „*Am Montag siebenundzwanzigsten Juli 2019 haben Siebzehntausend Neuhundert Achtundachtzig Nutzer aus Deutschland mit mobilen Endgeräten. Das sind acht Prozent weniger als am Montag davor*“, als nach diesen Informationen irgendwo in den Tiefen des Reportingsystems mit Mausklicks zu suchen. Diese Einfachheit entsteht durch beide Elemente des Vorgangs: Die Möglichkeit der Äußerung einer Aufgabe und die Fähigkeit des Systems, den Auftrag zu erkennen und darauf angemessen zu reagieren.

11.6 Intelligente Assistenten, die etwas für Menschen erledigen

Das einfachste Beispiel für Agenten, die etwas für Menschen erledigen sind eindimensionale Vorgänge, wie das An- und Abschalten von Licht, das Starten von Musik, das Verändern von Lautstärke und Raumtemperatur. Die großen Plattformbetreiber wetteifern darin, ihre Assistenten möglichst breit zu lizenziieren und so bietet Ikea Lampen an, die mit Amazon Alexa fernzusteuern sind. Anbieter von Kühlschränken und Mikrowellenherden integrieren Apples Siri. Apple hat mit Apple Home Kit auch ein Angebot an Hersteller von Lichtschaltern, Türschlössern und vielen anderen Smart Home-Komponenten gemacht. Google hat sein Aquisition Nest nunmehr vollständig in das eigene Ökosystem integriert und die Geräte firmieren unter Nest Hub. Sie lassen die Heimsteuerung per Voice und Assistant zu – auch aus der Ferne.

Intelligente Assistenten erledigen für Nutzer aber auch teilweise komplexe Vorgänge, wie das Einlegen von Einsprüchen bei Verkehrsverstößen. Bereits 2016 war der DoNotPay-Bot von einem britischen Jurastudenten entwickelt worden. Mit dem damals in Facebook Messenger umgesetzten Bot konnte man unter Eingabe einiger Daten den optimierten

Einspruch gegen Verkehrsstrafen einlegen. In einem strukturierten Prozess führte der Bot den Nutzer durch jeden Schritt mit dem Erfolg, dass bereits nach den ersten Monaten mehr als 200.000 Menschen mithilfe von DoNotPay eingestellte Verfahren vom Betreiber des eigentlich als Hobbyprojekt gestarteten Projekts gemeldet wurden. Das amerikanische Unternehmen [LegalShield.com](#) bietet seinen Kunden einen vergleichbaren Bot an, der neben Verkehrsstrafen noch weitere Rechtsgebiete abdeckt.

Erledigungsassistenten werden auch in innerbetrieblichen Fragen eingesetzt, so gibt es verschiedene Systeme, die typische Fragestellungen des Personalmanagements angehen und Mitarbeitern über einfache im Intranet installierte Chatbots die Möglichkeit zur Einreichung eines Urlaubsantrags oder der Krankmeldung ermöglichen. Damit machen sie die Suche nach den richtigen Antragsformularen im Intranet überflüssig und verbessern durch die im Bot integrierte Ausfüllhilfe auch die Qualität der Anträge.

11.7 Bots, die Inhalte zusammentragen und vorlegen

Vom Bank-One-Agenten Eni, der Bankkunden bei bestimmten Ereignissen benachrichtigt und aufmerksam macht, war bereits die Rede und davon, dass diese Art von Service auch im professionellen Umfeld an unterschiedlichen Stellen eingesetzt werden kann. Es gibt mehrere intelligente Softwareagenten, die für Systemadministratoren automatisiert Statistiken der von ihnen verantworteten Server und Systeme zusammentragen und auf besondere Aspekte hinweisen.

Es gibt eine große Zahl von Slack-Bots, die Softwareentwickler bei Ihrer Arbeit unterstützen. So zum Beispiel das indisches-amerikanische Startup [YellowAnt.com](#), das auf den natürlichsprachigen Textbefehl: „YellowAnt – trage mir die Bilder für das Projekt xyz aus der Dropbox und die Codesnippets aus Github zusammen und zeige mir das Briefing“, alle notwendigen Bestandteile für die Bearbeitung des betreffenden Projekts durch einen Programmierer an einen benannten Ort kopiert und sich nach erfolgreicher Erledigung beim Auftraggeber per Slack-Nachricht meldet. Informationsbeschaffung über alle Hürden der hierarchischen Navigationsstrukturen hinweg ist eine besondere Spezialität von intelligenten Assistenten, nicht nur im beruflichen Umfeld von Programmierern, sondern vor allem natürlich für Verbraucher.

Verschiedene Anbieter versprechen einfache Tools mit Hilfe derer wenig genutzte, weil schwer zugängliche FAQs, die auf vielen Webseiten ein trauriges Schattendasein fristen, obwohl die in ihnen gefassten Informationen für Nutzer sehr wertvoll sind, schnell in für Bots zugängliche Formate umformatiert werden können. Am Ende einer solchen Bemühung steht ein „Chatbot als Betriebsanleitung“. Statt Informationen suchen – Informationen finden, unterstützt von einem Softwareagenten, der Informationen zusammenträgt, strukturiert und dem Nutzer im Sinne einer Empfehlung kontextualisiert vorlegt.

Automatisierte Empfehlungen sind vielen von uns eher auf ungute Weise bekannt. Amazon-Recommendations und Facebook-Anzeigen, die auf schmerzhafte Weise am tatsächlichen Interesse und Bedarf vorbeiziehen: Targeting aus der Hölle. Durch CUIs und intelligente Assistenten besteht Hoffnung, dass die hinterlegten Algorithmen nicht mehr

ausschließlich auf Datenabgleich der vermeintlich eigenen Bedürfnisse mit den historischen Bedürfnissen vergleichbarer Nutzer fußen, sondern im Dialog in Echtzeit mit dem Nutzer geformt werden. Conversational Commerce verspricht die tatsächlichen Bedürfnisse von Nutzern im Dialog zu erfassen und gute Empfehlungen mit dem System zu schärfen: „Chatbot statt Suchfeld“.

Bei all diesen Anwendungsfeldern stellt sich immer wieder die Frage: Wie können Wissen und Erfahrung zusammengetragen und in möglichst optimierter Weise aufbereitet werden? Sei es in Form von reduzierten Voice-Nachrichten oder in Form von ebenfalls sparsamen visuellen Informationen. Knappe Snippets in Themengebieten wie Reisen, Navigation usw., wie man sie in modernen Suchmaschinen derzeit vorfindet, weisen hier den Weg auch für die Art, wie Inhalte aggregiert und in intelligenten Assistenten künftig verhandelt werden.

Ein gutes Beispiel, mit wie viel Liebe zum Detail und der themenspezifischen Tonalität gearbeitet werden kann, zeigt der Skill „A Rooki's Guide to the NFL“. NFL-Fans erhalten über das Alexa Voiceinterface Zugang zu allen denkbaren Inhalten der NFL-Saison 2019. Frag Alexa-NFL und Du bekommst eine informative, unterhaltsame Antwort: Fans, die die Qualifikation aktivieren, haben Zugang zu mehr als 1000 Begriffen aus dem Bereich American Football und NFL, die sich über Bereiche wie Regeln, Positionen, Formationen, Ausrüstung, Spieler und Schlüsselpersonen erstrecken. Dieser Aspekt der Fertigkeit zielt eher darauf ab, Neueinsteiger mit Football-Begriffen wie „Pistole“, „Screen Pass“, „Nikel“ und anderen vertraut zu machen.

Fans können sich auch nach allgemeinen Informationen über Spieler erkundigen, wie z. B. „Wer ist Tom Brady?“ oder „Wo ist Lamar Jackson aufs College gegangen?“ oder „Wie groß ist Russell Wilson?“. Und sie können nach Spielplänen, Matches, Spielzeiten, TV-Netzwerk, Punktzahlen sowie Informationen über die Stadien fragen, welche Teams in einer bestimmten Konferenz oder Abteilung sind, wer die Cheftrainer sind und vieles mehr.

Der NFL-Skill ist auch in der Lage, sich an die Geschichte der Super Bowl zu erinnern und bietet die Punktzahl, den Ort und das Datum einer der letzten 52 Super Bowls, sowie den Super Bowl MVP und den Halbzeit-Act aus jedem Spiel.

Bei der Beantwortung von Fragen kann der Skill Antworten in einer Vielzahl von Formen zurückgeben, sowohl als kurze und längere („Go Long“) Definitionen, als auch als Videos und Bilder auf Alexa-Geräten mit Bildschirmen.

11.8 Bots, die Menschen prüfen und Sachverhalte erfassen

Conversational Interfaces werden in verschiedenen Zusammenhängen dafür eingesetzt, Sachverhalte zu erfassen und gleichzeitig die Äußerungen der Nutzer zu überprüfen und in den richtigen Kontext zu bringen.

Eine vergleichsweise einfache Anwendung ist die Transkription von Gesprochenem. Der indische Motorradhersteller TVS testet Systeme, die die Auftragsannahme in Werkstätten automatisch „mitschneiden“ und die beauftragten Arbeiten damit zweifelsfrei dokumentieren.

Das amerikanische Unternehmen AG Voice Global ist darauf spezialisiert, Inspektoren von Agrarbetrieben, die auf dem Feld die Entwicklung von Pflanzen dokumentieren, so zu unterstützen, dass die Erfassung von Zuständen oder Problemen automatisch und gut strukturiert in Echtzeit erfolgt, während der Inspektor seine Befunde ins Gerät spricht. Da auf den Anbauflächen nicht überall ausreichende Netzqualität vorhanden ist, wird diese Niederschrift on-device im Gerät durchgeführt. Sollten bestimmte geforderte Aspekte im Bericht zunächst nicht enthalten sein, fragt das System beim Inspektor nach und vervollständigt so den Bericht.

Ebenfalls AG Voice Global bietet Systeme für Tierärzte, die es den Veterinären ermöglichen, während der Behandlung selbst mittels Sprache einen Bericht über Behandlungsmethoden und eingesetzte Medikamente zu erstellen. In der Tiermedizin bestehen teilweise sehr hohe Dokumentationsanforderungen, denen auf diese Weise praktisch nebenbei Genüge getan werden kann.

Nicht nur in der Tiermedizin, sondern auch in der Humanmedizin gibt es hohe Erwartungen an sprachgesteuerte intelligente Assistenzsysteme (siehe Abb. 11.2). Chancen werden in allen Prozessschritten gesehen. Von der Vereinfachung von Verwaltungsaufgaben, zum Beispiel der Schritt-für-Schritt-Vereinbarung von Arztterminen, oder bei der Umsetzung der bereits erwähnten Dokumentationspflichten der Ärzte, die diese heute häufig eher auf den Monitor blicken lassen, als in die Augen des vor ihm sitzenden Patienten. Intelligente Assistenten können auch vorbeugend wertvolle Aufgaben übernehmen, indem Patienten im Dialog zu Fitness und gesunder Ernährung angeleitet werden. In der Diagnosephase können intelligente Assistenten individuelle Probleme und Konstellationen in einer Gesprächssituation genau erfassen. Auch in der eigentlichen Behandlung spielen intelligente Assistenten eine Rolle. Insbesondere im Bereich der psychischen Erkrankungen und

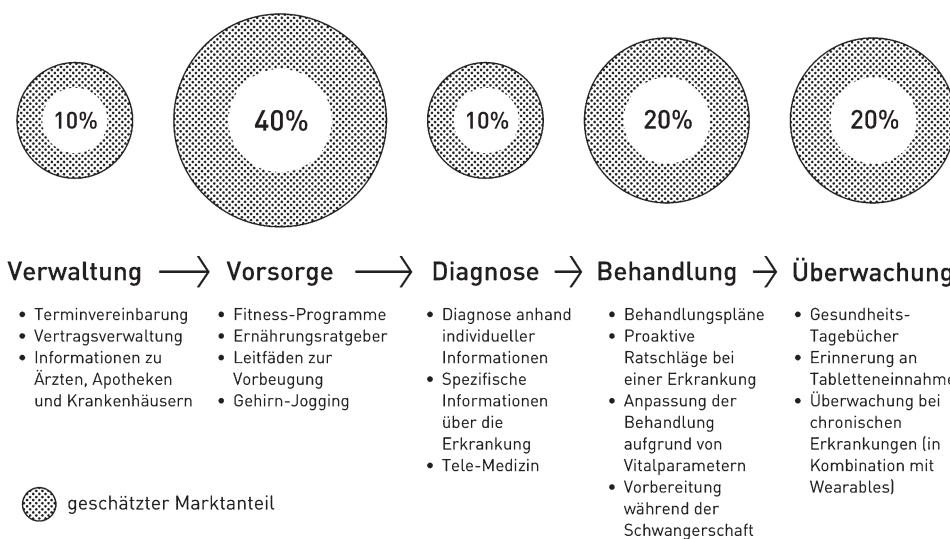


Abb. 11.2 Intelligente Assistenten im Gesundheitssektor

Probleme werden Sprachassistenten heute bereits eingesetzt. Schließlich wird auch das Monitoring von intelligenten Assistenten unterstützt, die den Zustand und Heilungsfortschritt regelmäßig abfragen sowie Erinnerungen an bestimmte Therapieschritte vornehmen können.

Beispieldialog

- CUI: Können Sie irgendeine Veränderung in Ihrem Nutzungsverhalten erkennen?
- Anwender: Ja, ich habe den Eindruck viel besser zu schlafen und mich im Tagesverlauf auch mehr zu bewegen.
- CUI: Sehr gute Beobachtung! Tatsächlich verbessert sich Ihre Schlafqualität um zehn Prozent, wenn Sie täglich mehr als 1000 Schritte machen. Konnten Sie noch etwas anderes beobachten?
- Anwender: Ich glaube nicht. Habe ich etwas übersehen?
- CUI: Also ich habe ebenfalls festgestellt, dass Sie an diesem Wochenende deutlich mehr gelaufen sind. Ist das wichtig für Sie?
- Anwender: Interessant, können Sie mir weitere Informationen geben?
- CUI: Sicher. An den vergangenen Wochenenden sind Sie täglich jeweils nur 3000 Schritte gegangen. Am letzten Wochenende dagegen 5000 Schritte. Soll ich die Art ihrer Bewegung weiterhin aufzeichnen und Sie auf Verhaltensänderungen aufmerksam machen?

Es stellt sich die Frage, ob Maschinen genügend Empathie entwickeln können, um Menschen bei wichtigen Fragen zur körperlichen und mentalen Gesundheit begleiten zu können, oder ob die Anonymität der Maschine, das Unpersönliche, nicht hinderlich ist.

Es zeigt sich allerdings, dass es eine ganze Reihe von Anwendungsfällen gibt, bei denen die Anonymität kein Problem, sondern wie in der echten Welt, eine Chance darstellt. In Brasilien betreiben die Anonymen Alkoholiker einen Bot [5], der Alkoholranke wie ein alter Freund betreut – offensichtlich mit sehr guten Ergebnissen.

Einen etwas anderen Ansatz verfolgt das amerikanische Unternehmen x2ai.com, das mit „Tess“ einen Bot betreibt, der von klinischen Psychologen erarbeitet wurde. Tess ist ein Mental Health Chatbot, der Menschen durch schwierige Zeiten coacht, um Widerstandsfähigkeit aufzubauen, indem er natürlichsprachige Gespräche via Chat führt – ähnlich wie beim Chatten mit einem Freund oder Coach – und dabei hervorragende Ergebnisse bei der Behandlung von Depressionen und Angststörungen erzielt. Grundlage dieser Erfolge ist offenbar die ständige Verfügbarkeit. Denn dadurch wird das Problem gelöst, dass Störungen in der Regel ja nicht während des eigentlichen Gesprächs mit dem Therapeuten behoben werden, sondern in den teilweise langen Phasen dazwischen. Tess wird in den USA kommerziell vermarktet. Es wurde aber auch bereits im Frühjahr 2016 in England versuchsweise bei der Betreuung von syrischen Kriegsflüchtlingen [7] mit ebenfalls guten Erfolgen eingesetzt.

Coaching und psychologische Begleitung mithilfe von intelligenten Assistenten wird von vielen Anbietern erprobt: Replika.ai ist ein psychologischer Begleiter, entwickelt in den USA. Replika analysiert die Texteingaben der Nutzer und reagiert darauf. Das amerikanische Magazin Wired [8] schreibt im Dezember 2017: „Die Verwendung von Replika kann sich auch therapeutisch anfühlen, in gewisser Weise. Die App bietet einen Raum, um sich ohne Schuldgefühle zu erleichtern, über komplizierte Gefühle zu sprechen, alle eigenen Gedanken ohne Urteil zu äußern.“

Die Möglichkeit, sich ohne Angst vor Repressalien äußern zu können, wird auch bei Ansätzen von Unternehmen in den USA, Kanada und Europa genutzt, bei denen es um die Meldung von sexuellen Übergriffen in Unternehmen geht. „Spot“ beispielsweise bietet Mitarbeitern die Möglichkeit, Belästigungen in einem Conversational Interface zu melden und dabei einerseits keine Sorge vor Repressalien haben zu müssen. Andererseits wird dadurch auch sichergestellt, dass die Aussagen bei einer Befragung der Betroffenen durch einen echten Menschen durch Voreingenommenheit des Fragenden oder ungeeignete Fragenformulierung verfälscht werden – und in einem späteren Verfahren dadurch untauglich sind. Durch die Erfassung in Bot-Systemen [9] wie „Spot“ wird die Befragung automatisch auch ordnungsgemäß mit Zeitstempel dokumentiert und steht für eventuell notwendige weitere Schritte zur Verfügung.

Ein weiterer Vorteil der Erfassung durch Bots gegenüber der Erfassung durch Menschen ist die Qualität der Dokumentation, auf die im Bereich von Kreditanträgen oder Beratungsangeboten durch Banken gesetzt wird. Wurde der Kunde ausreichend auf die Risiken eines Finanzproduktes informiert? Hat er durch entsprechende Nachfragen gezeigt, dass er die Funktionsweisen des Produkts verstanden hat? Hat der Kunde die Umstände bei einem Kreditantrag umfassend und vollständig geschildert? All diese Fragen führen in der Praxis immer wieder zu unguten Überraschungen bei Finanzinstituten und ihren Kunden. Diese können durch den Einsatz von Chatbots und Voice-Assistants ausgeschlossen werden, weshalb die Finanzindustrie so intensiv wie sonst keine Branche mit dem möglichen Einsatz von CUIs beschäftigt ist. So experimentiert die Deutsche Kreditbank AG (DKB) in Kooperation mit dem Fintech-Unternehmen Finreach seit März 2017 und setzt einen Roboter für Kommunikationsaufgaben ein. Der „Herbie“ getaufte Chatbot wird Kunden der BayernLB-Tochter DKB auf der Website zunächst zu einem Konsumentenkredit von bis zu 50.000 Euro begleiten und alle Schritte im Rahmen des Kreditantrags im Dialog durchführen.

Im Feld von Human Ressources, insbesondere im Recruiting, gibt es ebenfalls eine Reihe von Ansätzen, die Qualität von Bewerbungsgesprächen durch den Einsatz von intelligenten Assistenten zu verbessern. Einerseits um Antworten der Bewerber besser dokumentieren zu können, andererseits aber auch, um die mögliche Voreingenommenheit seitens der Recruiter gegenüber Geschlecht, Herkunft und Aussehen zu minimieren. Die schwedische Firma Furhat Robotics, hat einen Bot namens Tengai getestet und entwickelt, der automatisiert Jobinterviews führt. Dabei ist die Anordnung so gewählt, dass der Bot dem Bewerber wie ein Roboter gegenübersteht. Einmal abgesehen von der Frage, ob der-

artige Anthropomorphismen – also die Nachbildung menschlicher Erscheinungen – tatsächlich sinnvoll sind, oder nicht eher abschrecken, ist die Absicht die Erfassung und Prüfung durch Assistenten statt durch Menschen durchführen zu lassen, in vielen Anwendungsbereichen recht erfolgversprechend.

11.9 Assistenten, die zwischen Menschen vermitteln

Aktuell stehen Anwendungen, bei denen intelligente Assistenten zwischen zwei oder mehr Menschen vermitteln, noch sehr am Anfang. Im Rahmen von FAQ-Bots können Kunden schneller zum richtigen Ansprechpartner innerhalb eines Unternehmens finden.

Über die strukturierte Erfassung von Umständen und Bedürfnissen gibt es erste Versuche, Gärtner, Klavierlehrer und ähnliche persönliche Unterstützer mittels Software-Agenten aus den Angebotsdatenbanken zu filtern und anschließend zwischen Nachfrager und Anbieter einen Kennenlerntermin zu organisieren und jedem der Beteiligten, gewissermaßen als Gesprächsvorbereitung, alle Eigenschaften des jeweils anderen, die im Dialog mit dem Software-Agenten erfasst wurden, zur Verfügung zu stellen.

Ein Beispiel wie Assistenten Arbeitsabläufe verändern können, ist AIVA, ein Software-System, das mit Amazon Alexa oder Google Assistant Krankenhauspatienten oder Bewohnern von Seniorenheimen die Möglichkeit gibt, per Voice oder durch Nutzung der zu AIVA gehörigen Mobile-App, ihre Bedürfnisse zu äußern, zum Beispiel den Wunsch auf Toilette zu gehen. Diese Anfrage wird den verantwortlichen Pflegern in einer App gemeldet und abhängig von der Dringlichkeit in eine Auftragsliste geschrieben, die die Pfleger dann abarbeiten. Das Versprechen von AIVA: Wir sind der weltweit erste Voice-Pflege-Assistant, der Hands-free-Kommunikation für zufriedenere Patienten anbietet und gleichzeitig den Workflow von Pflegern optimiert. Das System wird derzeit im Cedars-Sinai-Hospital in New York in mehreren hundert Krankenzimmern getestet.

Ein weiteres spektakuläres Beispiel für Assistenten, die Menschen verbinden ist Google Duplex. Ein System, von dem in diesem Buch immer wieder die Rede ist, da es seit seiner Vorstellung im Mai 2018 Fachleute beeindruckt und immer wieder in der Presse zitiert wird. Allerdings muss man anmerken, dass Google Duplex zum Zeitpunkt der Veröffentlichung dieses Buches noch nicht kommerziell verfügbar ist. Die Vision dahinter, dass ein Software-Agent eigenständig bei einem Friseur anruft und autonom einen Termin vereinbart, zeigt, dass intelligente Assistenten, die zwischen Menschen vermitteln, möglicherweise in nicht allzu ferner Zukunft ein sehr wichtiges Aufgabenfeld übernehmen können. Wenn man sich vergegenwärtigt, wie disruptiv eine solche neu zur Verfügung stehende Vermittlung zwischen Menschen sein kann, zeigt das Unternehmen Uber, das mit einer zunächst relativ einfachen App ja auch nichts anderes anbietet, als Fahrer und Fahrgäste zum richtigen Zeitpunkt auf bequeme Weise zusammen zu bringen.

11.10 Bots, die aktiv in Dialog gehen

Intelligente Assistenten müssen nicht ausschließlich auf den Befehl von Menschen warten, bis sie aktiv werden. In unterschiedlichen Anwendungsgebieten nehmen Bots eine aktiver Rolle ein. Dies gilt beispielsweise bei Sprach-Lern-Assistenten, wie der in Kap. 2 beschriebenen Lily, die virtuelle Sprach-Lehrerin, die mit ihren Nutzern einen aktiven Dialog führt. Im Bereich der Pflege von Alten und Kranken existieren neben dem in Kap. 6 beschriebenen „Elliq – der Sidekick für glücklicheres Altern“, eine ganze Reihe von weiteren Angeboten und Versuchen, die teilweise deutlich weniger ambitioniert sind als Eliiq, das neben dem digitalen Assistant auch eine eigene Hardware hat und durch Einsatz von Biometrie aus den Stimmen der Nutzer Rückschlüsse auf den aktuellen Gemütszustand zieht. Immer geht es darum, Menschen in Dialoge zu verwickeln und damit die geistige Beweglichkeit und Zufriedenheit zu steigern. In diesen Bereich fallen auch viele Spiele und Quizzes, die Menschen mit intelligenten Assistenten spielen, um sich zu unterhalten.

Bei Otto und anderen E-Commerce-Unternehmen werden Versuche unternommen, Kunden gleich beim Auspacken der Lieferung in einen Dialog mit einem Voice-Assistenten zu bringen und auf diese Weise die Zufriedenheit des Kunden mit dem Kauf abzufragen und eventuelle Probleme – beispielsweise mit der Größe – sofort zu erkennen und ggf. lösen zu können.

In der englischen TV-Serie „Black Mirror“ [10] geht es um die dunklen Aspekte der modernen Gesellschaft, insbesondere in Bezug auf unser Verhältnis zur Technologie. Jede Episode zeigt ein Bild einer Welt, die futuristisch und doch glaubwürdig ist, cool und doch auch erschreckend. In der Folge „Bin gleich zurück“ erzählt die Geschichte bereits 2013 von einer Frau, deren telefonbesessener Freund plötzlich stirbt. Die Frau wird mit einem experimentellen Dienst bekannt gemacht, der die Lebenden wieder mit den Toten verbindet. Mit Hilfe von Fotos, Videos und anderen online verfügbaren Daten bietet der Dienst den Lebenden die Möglichkeit, mit ihren verstorbenen Angehörigen zu sprechen. Je mehr Daten Sie eingeben, umso genauer ist die Version des Dienstes von ihrem verstorbenen Angehörigen. Es beginnt mit geschriebenen Texten, in denen ein Bot die Schreibgewohnheiten und den Humor des Verstorbenen nachahmt. Die nächste Stufe sind Stimmabnahmen, die die Stimme des verstorbenen Menschen in Anrufen imitieren. Und dann gibt es noch das letzte Level: ein Klon des verstorbenen Geliebten. Der Kern von all dem sind Daten. In diesem Fall ist fast alles, was online existiert, nützlich, von E-Mails bis zu Heimvideos. Die Version ihres verstorbenen Angehörigen, die der Dienst zur Verfügung stellt (mit ein wenig Spielraum für Anpassungen), basiert vollständig auf diesen Daten. Wer wir heute online sind, wird bestimmen, wie wir in Erinnerung bleiben. Faszinierend und ziemlich gruselig zugleich. Sicherlich eine ferne Zukunftsvision, die allerdings möglicherweise bereits näher ist, als manche dies glauben mögen.

11.11 Aus Wearables werden Awareables

Unsere Geräte werden uns bald besser kennen als wir unsere Geräte. Das Smartphone enthält wahrscheinlich mehr Sensoren als jedes andere Gerät in unserem Zuhause. Die Wearables werden mitfühlend und aufmerksam über ihren Träger „wachen“. Doch fortschrittliche Technologien wie die Gesichtserkennung werden derzeit nur für scheinbar einfache Dinge wie die Freischaltung des Telefons eingesetzt. Aber was ist, wenn Ihr Smartphone sich Ihrer voll bewusst wird; nicht nur was Sie tun, sondern auch wer Sie sind? Heute können kognitive Dienste den Gemütszustand einer Person bereits „verstehen“, wenn sie ihnen nur in die Augen [11] schauen oder wenn sie die Stimme hören. In der bereits mehrfach durchgeführten qualitativen Studie von Ericsson gab Ende 2018 die Hälfte derjenigen Befragten, die derzeit virtuelle Assistenten (Google Now, Alexa oder Siri, etc.) in Handys oder Smartspeakern verwenden, an, dass sie glauben, dass Smartphones ihre Emotionen bald verstehen werden. Und 42 Prozent glauben, dass Smartphones sie bald besser verstehen werden als ihre Freunde. Tatsächlich denken 6 von 10 der Befragten, dass Geräte, die unsere Stimmungen wahrnehmen und darauf reagieren, in drei Jahren Mainstream sein werden. In diesem Fall reagieren die Geräte eher auf Situationen als auf Befehle.

Stellen Sie sich vor, Sie hätten Ihrem intelligenten Assistenten ein Geheimnis verraten und wollten, dass er diskret ist. Sechsundfünfzig Prozent der Benutzer von intelligenten Assistenten wünschen sich einen Lautsprecher, der in einer solchen Situation seine Stimme reduziert. Ebenso wünschen sich 52 Prozent eine Smartwatch, die Benachrichtigungen blockiert, wenn sie versteht, dass Sie beschäftigt sind. Siebenundvierzig Prozent wollen eine App, die einen Social-Media-Beitrag überprüft, nur weil sie einen überraschten Gesichtsausdruck bemerken. Wenn Geräte uns so gut kennen, werden sie auch unsere Geheimnisse kennen. Aber erstaunlicherweise vertrauen doppelt so viele einem Gerät mehr als einem Menschen, wenn es darum geht, ihre Geheimnisse zu bewahren.

Im 2014 veröffentlichten amerikanischen Spielfilm „Her“ verliebt sich der frisch getrennte Theodore – gespielt von Joaquin Phoenix – in einen offenbar weiblichen virtuellen Assistenten namens Samantha. Der Film spielt in naher Zukunft in Los Angeles. Wenn Theodore das System startet, lässt er sich von Samanthas freundlicher Stimme bezaubern, denn sie stellt sich mit viel Verständnis, Sensibilität und erstaunlichem Humor auf ihn ein. Samanthas und Theodores Bedürfnisse und Sehnsüchte ergänzen sich und entwickeln sich weiter. Ihre Freundschaft wird intensiver bis sie sich schließlich ineinander verlieben. Wenn zutrifft, dass Science Fiction häufig die tatsächliche Zukunft voraus nimmt und damit auch den weiteren Verlauf der technischen Entwicklung beeinflusst, dann stehen wir – wie Theodore in Her – angesichts der weiteren Entwicklung von Conversational User Interfaces und intelligenten Assistenten künftig vor einer ganzen Reihe von Herausforderungen. Die Tageszeitung Die Welt nannte den Film „*„betörend schön, formal brillant, eine philosophische Herausforderung.“ Ein guter Ausblick in unsere Zukunft* [12].“

Literatur

1. Apple: Introducing Voice Control on Mac and iOS — Apple, (04.06.2019), <https://www.youtube.com/watch?v=aqoXFCCTfm4&feature=youtu.be>, zugegriffen 20.11.2019
2. Patentscope: Title of Invention : SILENT VOICE INPUT, <https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2018223388&tab=FULLTEXT&maxRec=1000>, zugegriffen 20.11.2019
3. Oracle: Oracle Unveils AI-Voice for the Enterprise, (17.09.2019), <https://www.oracle.com/corporate/pressrelease/oow19-oracle-unveils-ai-voice-for-enterprise-091719.html>, zugegriffen 03.11.2019
4. Capital One: Eno, <https://www.capitalone.com/applications/eno/>, zugegriffen 12.11.2019
5. Alcoholics Anonymous: Reaching new audiences with a bot for Messenger, <https://www.facebook.com/business/success/alcoolicos-anonimos>, zugegriffen 20.11.2019
7. Solon, Olivia: Karim the AI delivers psychological support to Syrian refugees, (22.3.2016), <https://www.theguardian.com/technology/2016/mar/22/karim-the-ai-delivers-psychological-support-to-syrian-refugees>, zugegriffen 30.11.2019
8. Pardes, Arielle: What My Personal Chat Bot Is Teaching Me About AI's Future, (11.12.2017), [https://www.wired.com/story/what-my-personal-chat-bot-replika-is-teaching-me-about-artificial-intelligence/amp](https://www.wired.com/story/what-my-personal-chat-bot-replika-is-teaching-me-about-artificial-intelligence/), zugegriffen 30.11.2019
9. Johnson, Khari: Spot launches workplace harassment bot for internal company use, (23.10.2018), <https://venturebeat.com/2018/10/23/spot-launches-workplace-harassment-bot-for-internal-company-use/>, zugegriffen 30.11.2019
10. IMDB: Be Right Back, (21.10.2016), <https://www.imdb.com/title/tt2290780/>, zugegriffen 30.11.2019
11. Hoppe, Sabrina et al: Eye Movements During Everyday Behavior Predict Personality Traits, (13.04.2018), <https://www.frontiersin.org/articles/10.3389/fnhum.2018.00105/full>, zugegriffen 23.11.2019
12. Rodek, Hanns-Georg: Vom Zauber der Liebe zu einer Computerstimme (26.03.2014) <https://www.welt.de/kultur/kino/article126211315/Vom-Zauber-der-Liebe-zu-einer-Computerstimme.html>, zugegriffen 12.02.2020



Machen: Der Friesenbot

12

Zusammenfassung

Nach der Theorie kommt die Praxis. Wenn Sie nun neugierig auf Ihren eigenen Assistenten sind, haben wir hier die Anleitung dazu. Auf der Plattform Dialogflow kann man auch ohne Programmierkenntnisse einen sprachgesteuerten Assistenten zum Leben erwecken. Eine Anleitung für den Friesenbot.

Bekanntermaßen geht nichts über die direkte Erfahrung. Daher wollen wir an dieser Stelle ein kleines Tutorial anbieten, das es auch vollkommen unerfahrenen Lesern und Menschen ohne ausgeprägte technische Fähigkeiten ermöglicht, innerhalb kurzer Zeit einen kleinen eigenen Chatbot zu bauen, der sofort funktioniert und über Smartphone und Smartspeaker nutzbar ist.

Auf diese Weise versteht jeder, was die praktischen Grundprinzipien von Conversational User Interfaces sind und auch, wie der aktuelle Stand von Tools und Technologien ist, die im Netz für Jedermann frei angeboten werden. Wir haben uns dabei für die Google Platform entschieden, da diese recht offen ist und es damit möglich ist, auch andere, weitere Dienste einzubinden. Zudem kann man damit das eigene CUI auch auf jedem Smartphone unterwegs ausprobieren.

In diesem Tutorial werden Sie den Friesenbot bauen. Eine Google Action, die den Nutzern Witze erzählen kann und ihnen außerdem das Wetter mitteilt. Dabei beginnen Sie mit den Basics: Wie befülle ich einen Intent? Und wie kann ein Nutzer diesen aufrufen? Dann führen wir Sie hin zu komplexeren Bereichen wie dem Anbinden einer API, um das echte Wetter live abzufragen. Das Bauen dieser Action gibt Ihnen einen groben Überblick darüber, wie man ganz praktisch einen digitalen Assistenten baut.

Lassen Sie uns starten

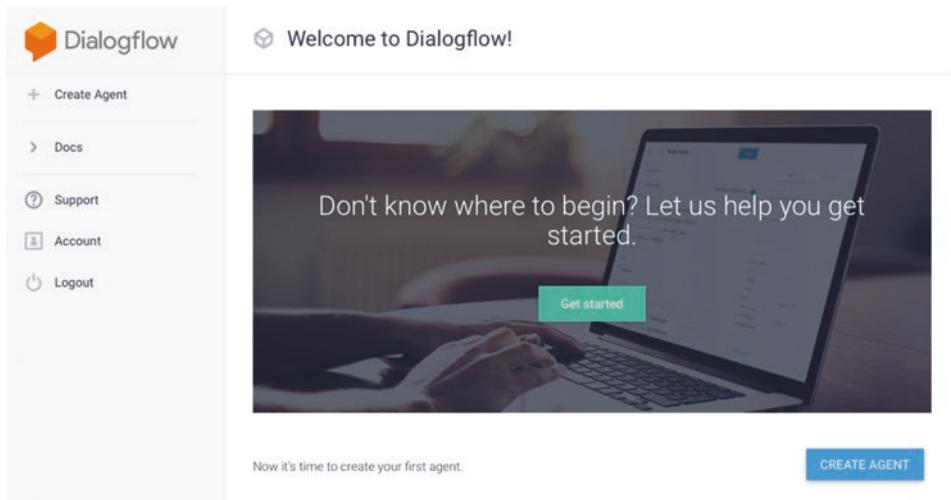
- Gehen Sie auf www.dialogflow.com und klicken Sie dort auf die Schaltfläche „Sign In“ um sich anzumelden.

The screenshot shows the official Dialogflow website. At the top, there's a navigation bar with links for Overview, Case studies, Docs, Blog, Pricing, and Support. To the right of the navigation is a search bar and links for Go to console and Sign in. Below the navigation, a large banner features the text "Build natural and rich conversational experiences" and "Intro to Dialogflow". It includes a video thumbnail titled "Intro to Dialogflow" with options to "Watch later" and "Share". A woman is shown sitting at a desk with a laptop displaying the Dialogflow logo. Below the banner, there are three columns of text: "Powered by Google's machine learning", "Built on Google infrastructure", and "Optimized for the Google Assistant". Each column has a brief description and a "Sign up for free" button. At the bottom, a teal-colored box contains a testimonial from Mandi Galluch: "Dialogflow is user-friendly, intuitive, and just makes sense. Its natural language processing (NLP) is the best we've tried." followed by her title "Mandi Galluch, Digital Experience Program Leader, Domino's".

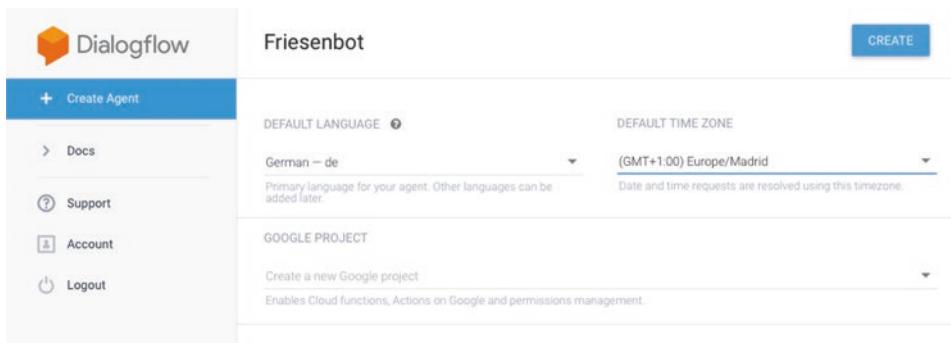
- Melden Sie sich mit Ihrem Google Account bei Dialogflow.com an.



- Klicken Sie auf den Menüpunkt „Create Agent“ um einen Agenten zu erstellen.



- Geben Sie Ihrem Agenten einen Namen, in unserem Beispiel „Friesenbot“, und wählen Sie die gewünschte Sprache und Zeitzone aus.



- Zunächst erstellen Sie den sogenannten „Welcome-Intent“. Klicken Sie hierfür auf die Schaltfläche „Default Welcome Intent“. Der „Welcome-Intent“ ist die Begrüßung Ihres Agenten und wird bei jedem Start der Interaktion aufgerufen. An dieser Stelle veranschaulichen Sie den Usern die Funktionsweise Ihres Agenten.

The screenshot shows the Dialogflow web interface. On the left, there's a sidebar with tabs for 'Intents', 'Entities', and 'Knowledge [beta]'. The main area is titled 'Intents' and contains a search bar and a list of intents. Two intents are listed: 'Default Fallback Intent' (indicated by a blue icon) and 'Default Welcome Intent' (indicated by a blue dot). At the top right of the main area, there's a 'CREATE INTENT' button.

- Unter dem Menüpunkt „Responses“ können Sie alle sich dort befindlichen Responses löschen, sodass nur die Aufforderung „Enter a text response variant“ übrig bleibt. Dort tragen Sie den Begrüßungstext Ihres Agenten ein.

In unserem Beispiel sollen die User erfahren, dass sie Wetterinformationen und Ostfriesenwitze erfragen können, unser „Welcome-Intent“ lautet daher: „Hallo ich bin der Friesenbot. Ich habe die besten Witze der ganzen Küste für dich, kann dir aber auch das Wetter verraten. Möchtest du einen Witz hören, oder das Wetter abfragen?“

Klicken Sie nach der erfolgreichen Eingabe auf „Save“ um Ihren „Welcome-Intent“ zu speichern.

The screenshot shows the configuration of the 'Default Welcome Intent' in Dialogflow. The sidebar on the left has a 'Validation' checkbox checked. The main area shows the intent configuration. Under 'Parameters', there are four columns: REQUIRED, PARAMETER NAME, ENTITY, and VALUE. The first row has 'Enter name' under 'PARAMETER NAME' and 'Enter entity' under 'ENTITY'. Under 'Responses', there are two sections: 'DEFAULT' and 'GOOGLE ASSISTANT'. The 'DEFAULT' section contains a 'Text Response' with two numbered steps: 1. 'Hallo ich bin der Friesenbot. Ich habe die besten Witze der ganzen Küste für dich, kann dir aber auch das Wetter verraten. Möchtest du einen Witz hören, oder das Wetter abfragen?' and 2. 'Enter a text response variant'. Below this is a 'ADD RESPONSES' button. At the bottom, there's a switch labeled 'Set this intent as end of conversation'.

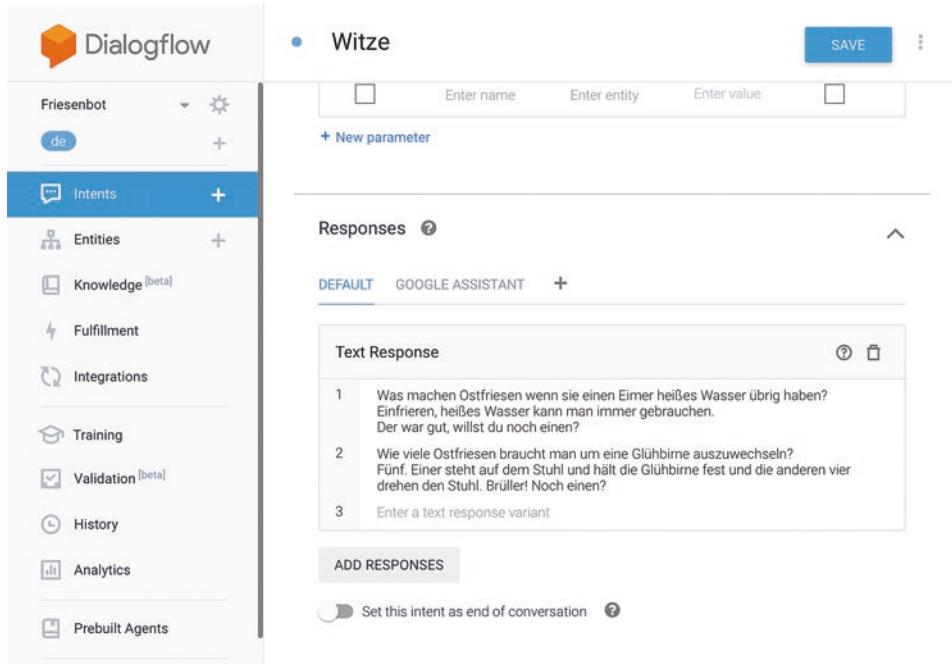
- Um unsere Änderung zu testen benutzen Sie die Konsole oben rechts. Dort können Sie die Funktionalität Ihres Agenten überprüfen. Geben Sie dazu z. B. „Hallo“ oder eine andere der „Trainings-Phrasen“ ein und bestätigen Sie Ihre Eingabe mit der Enter-Taste. Unter „DEFAULT RESPONSE“ sollten Sie die von Ihnen vorgegebene Antwort sehen, in diesem Fall Ihren „Welcome-Intent“. Geben Sie allerdings z. B. das Wort „Witze“ ein, wird der Agent mit einer Fehlermeldung antworten, da er noch nicht auf diese Phrase trainiert wurde.

- Um den Agenten auf diese Phrase zu trainieren, erstellen Sie einen neuen Intent mit dem Namen „Witze“. Unter „Trainingsphrasen“ geben wir dem Agenten Input, wie dieser Intent aufgerufen werden soll. Dafür können Sie Abwandlungen der Phrase „Witze“ wie z. B. „Witz bitte!“, „Einen Witz“ oder „Erzähl mir einen Witz!“ eintragen. Je mehr Abwandlungen Sie einfügen, umso besser kann Ihr Agent auf unterschiedlichen Input reagieren.

The screenshot shows the Dialogflow web interface for creating an intent named "Witze".

- Left Sidebar:** Shows navigation links for Friesenbot (selected), Entities, Knowledge [beta], Fulfilment, Integrations, Training, Validation [beta], History, Analytics, Prebuilt Agents, Docs (Standard Free), Support, Account, and Logout.
- Intent Overview:** Shows the intent name "Witze" with a "SAVE" button and a more options menu.
- Contexts:** A collapsed section.
- Events:** A collapsed section.
- Training phrases:** A list of user expressions:
 - 99 Add user expression
 - 99 Witz bitte
 - 99 Erzähl mir einen Witz
 - 99 Einen Witz!
 - 99 Ja
 A search bar "Search training phrases" is present above the list.
- Action and parameters:** A collapsed section.
- Extract the action and parameters:** A section containing a circular diagram with a central blue square and radiating lines, followed by text explaining parameters and a "MANAGE PARAMETERS AND ACTION" link.
- Responses:** A collapsed section.
- Fulfilment:** A collapsed section.

- Unter „Responses“ pflegen Sie nun die Antworten Ihres Agenten auf die vorher gegangenen Phrasen ein. Schreiben Sie dazu Ihre Lieblings-Ostfriesenwitze auf und fügen diese über „ADD RESPONSES“ ein. Auch hier gilt, je mehr desto besser. Testen Sie den Erfolg Ihres Inputs bei Bedarf erneut über die Konsole.



An dieser Stelle kann Ihr Agent bereits erfolgreich Witze erzählen. Im nächsten Schritt geht es darum den Agenten Informationen (in unserem Beispiel Wetterinformationen) aus dem Internet abrufen zu lassen

- Dazu erstellen Sie einen neuen Intent und benennen diesen „Wetter“. Unter die Trainingsphrasen schreiben Sie einige Phrasen wie z. B. „Wie wird das Wetter in Bremen?“ Oder „Wie ist das Wetter in Hamburg?“. Hamburg und Bremen werden mit farblichen Indikatoren durch Dialogflow markiert. Dies zeigt Ihnen, dass Dialogflow diese Städte als sogenannte „Variablen“ erkennt. Mit diesen Variablen können Sie in den „Text-Responses“ arbeiten, indem sie z. B. dort „Das Wetter in \$geo-city (Ihre Variable) wird heute schön.“ Speichern Sie den Intent und testen Sie diesen über die Konsole aus. Testen Sie dort andere Städte wie z. B. München werden Sie sehen, dass der Agent in seiner Antwort auch „Das Wetter in München wird heute schön“ ausgibt.

The screenshot shows the Dialogflow web interface for creating an intent named 'Wetter'. The left sidebar contains navigation links for Friesenbot, de, Intents (selected), Entities, Knowledge [beta], Fulfillment, Integrations, Training, Validation [beta] (checked), History, Analytics, Prebuilt Agents, Docs, Standard, and Upgrade. The main area is titled 'Wetter' with a 'SAVE' button. It includes sections for Contexts, Events, Training phrases, Action and parameters, and a preview pane.

Training phrases:

- Wie ist das Wetter in Hamburg?
- Wie ist das Wetter in **Hamburg**?
- Wie wird das Wetter in **Bremen**?

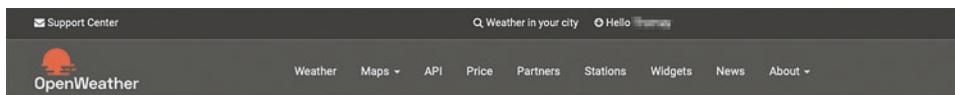
PARAMETER NAME	ENTITY	RESOLVED VALUE
geo-city	@sys.geo-city	Hamburg

Action and parameters:

Enter action name: (empty)

REQUIRED	PARAMETER NAME	ENTITY	VALUE	IS LIST
----------	----------------	--------	-------	---------

- Um eine konkrete Angabe der Wetterinformationen Ihres gewünschten Ortes zu erhalten, müssen Sie Ihren Agenten nun mit einer Wetter-API verbinden. In unserem Beispiel wird die API von openweathermap.org verwendet. Besuchen Sie die Seite und erstellen Sie sich dort einen kostenlosen Account.
- Klicken Sie, nachdem Sie sich erfolgreich eingeloggt haben auf Ihren Account und dann auf den Menüpunkt „API Keys“. Kopieren Sie sich nun den generierten API Key in ein leeres Dokument zur Zwischenablage. Die Freischaltung des API-Schlüssels kann einige Stunden dauern.



- Kehren Sie zurück in Ihren Intent bei Dialogflow. Aktivieren Sie dort im Untermenü „Fulfillment“ die Option „Enable webhook call for this intent“. Dies ermöglicht Ihrem Intent eine Informationsabfrage bei der API zu erstellen.

The screenshot shows the Dialogflow interface for the "Wetter" intent. On the left, there's a sidebar with navigation links: WetterBot, Intents (selected), Entities, Knowledge, Fulfillment, Integrations, Training, Validation (Beta), History, Analytics, Prebuilt Agents, and Docs. The main area shows the "Wetter" intent configuration. It includes sections for "Responses" (with a "Execute and respond to the user" section featuring a woman icon and a "ADD RESPONSE" button) and "Fulfillment" (with two checkboxes: "Enable webhook call for this intent" and "Enable webhook call for slot filling"). A "SAVE" button is located at the top right of the intent card. A "Try it now" button and a microphone icon are also present.

- Klicken Sie nun in der Seitenleiste auf „Fulfillment“ und aktivieren Sie dort die Funktion „Inline Editor“. Löschen Sie dann den gesamten Code in dem Fenster „index.js“ und fügen dort unseren bereitgestellten Code ein.

The screenshot shows the Dialogflow Fulfillment interface. On the left, there's a sidebar with various tabs: WetterBot, de, Intents, Entities, Knowledge [beta], Fulfillment (which is selected and highlighted in blue), Integrations, Training, Validation [beta], History, Analytics, Prebuilt Agents, Docs, Standard, and Upgrade. The main area is titled "Fulfillment" and has two sections: "Webhook" (disabled) and "Inline Editor (Powered by Cloud Functions for Firebase)" (enabled). The "Inline Editor" section contains a code editor with "index.js" and "package.json" tabs. The "index.js" tab shows the following code:

```

1 // See https://github.com/dialogflow/dialogflow-fulfillment-nodejs
2 // for Dialogflow fulfillment library docs, samples, and to report issues:
3 'use strict';
4
5 const axios = require('axios');
6
7 const functions = require('firebase-functions');
8 const {WebhookClient} = require('dialogflow-fulfillment');
9 const {Card, Suggestion} = require('dialogflow-fulfillment');
10
11 process.env.DEBUG = 'dialogflow:debug'; // enables lib debugging stats
12
13 exports.dialogflowFirebaseFulfillment = functions.https.onRequest((request, response) => {
14   const agent = new WebhookClient({ request, response });
15

```

Below the code editor, there are buttons for "View execution logs in the Firebase console" and "Last deployed on 10/15/2019 11:19". A "DEPLOY" button is also present. The status bar at the bottom indicates "Standard" and "Upgrade".

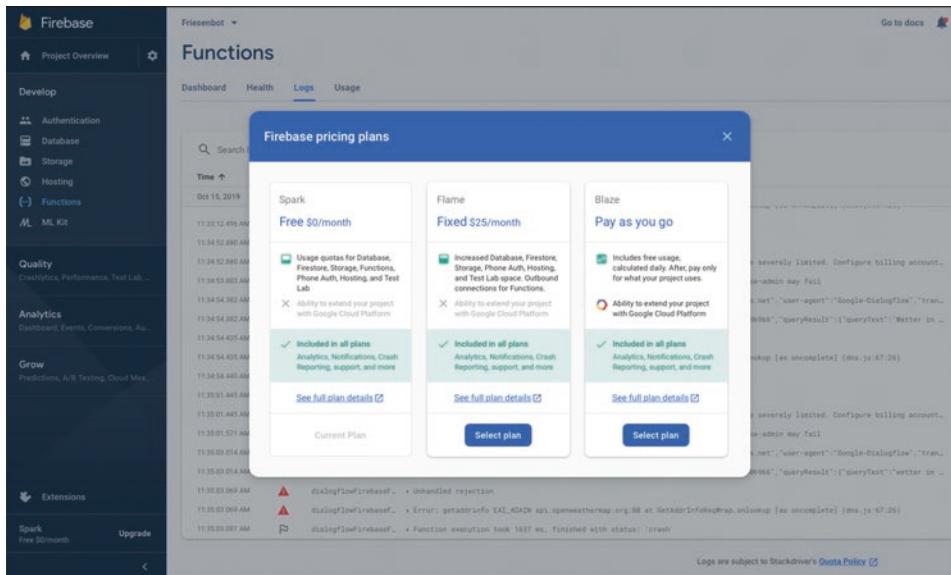
- Scrollen Sie zur Zeile 27 in unserem Code und fügen Sie bei „HIER API KEY EINSETZEN“ Ihren sich in der Zwischenablage befindlichen API-Schlüssel ein. Achten Sie darauf, nicht die Anführungszeichen zu löschen. Klicken Sie dann auf den Button „Deploy“ und warten Sie einen Moment. Ist der Prozess erfolgreich abgeschlossen, sollten Sie die Nachricht „Successfully deployed your Cloud Function and saved your Webhook“ lesen können.

```

26 function weatherHandler(agent){
27   const apiKey = 'HIER API KEY EINSETZEN';
28   const stadt = agent.parameters['geo-city'];
29   return axios.get(`http://api.openweathermap.org/data/2.5/weather?q=${stadt}&appid=${apiKey}&units=metric`)
30   .then((result) => {
31     agent.add(`Die Temperatur in ${stadt} liegt bei ` + result.data.main.temp + ` Grad Celsius`);

```

- Noch kann Ihr Agent Ihre Anfrage nicht erfolgreich bearbeiten. Dafür müssen Sie Ihren Account von „Free“ auf „Pay as you go“ upgraden. Hierbei entstehen aufgrund der niedrigen Zahl an Anfragen keine Kosten für Sie. Es ist trotzdem notwendig, Kreditkarten-Informationen zu hinterlegen, um im Prozess voranschreiten zu können. Neben dem Deploy Button lesen Sie „View execution logs in the Firebase console“. Nachdem Sie das Feld angeklickt haben, öffnet sich das „Firebase“-Menü. Klicken Sie dort auf „Upgrade“ und wählen Sie den Plan „Pay as you go“ aus. Geben Sie nun Ihre Kreditkarten-Informationen ein.



- Kehren Sie nun zu Ihrem Agenten zurück und testen Sie dessen Wetterinformationen einer spezifischen Stadt. Über die Konsole erhalten Sie passende Informationen. An diesem Punkt funktioniert Ihr Agent wie gewünscht und Sie können ihn mit gezieltem Testing und dem Erstellen weiterer Intents erweitern und optimieren.

The screenshot shows the Dialogflow Agent interface. At the top, there is a button labeled "Try it now" and a microphone icon. Below this, a link says "See how it works in Google Assistant, [↗](#)". The main area is titled "Agent".
USER SAYS: Wetter in Hamburg
COPY CURL: (button)
DEFAULT RESPONSE: Die Temperatur in Hamburg liegt bei 16.07 Grad Celsius
INTENT: Wetter
ACTION: *Not available*
PARAMETER VALUE
date
geo-city Hamburg
DIAGNOSTIC INFO

- Um ein besseres Verständnis für die im Hintergrund durchgeföhrten Abläufe zu erlangen können Sie sich den Code genauer anschauen. Ab diesem Punkt nehmen wir keine Veränderungen am Code vor. Gehen Sie zu Zeile 39 des Codes über den Menüpunkt „Fulfillment“ in der Seitenleiste. Dort verknüpfen wir unseren Intent „Wetter“ mit der Funktion, die dann ausgeführt werden soll. Wir nennen diese Funktion „WeatherHandler“. Der Rest des Codes ist Teil des Dialogflow-Systems und für uns in diesem Moment unwichtig.

```

35 // Run the proper function handler based on the matched Dialogflow intent name
36 let intentMap = new Map();
37 intentMap.set('Default Welcome Intent', welcome);
38 intentMap.set('Default Fallback Intent', fallback);
39 intentMap.set('Wetter', weatherHandler);
40 agent.handleRequest(intentMap);
41 });

```

- In Zeile 5 befindet sich „axios“, dessen Funktion das Anfragen der Informationen bei der von Ihnen eingebundenen Wetter-API ist.

```

4
5 const axios = require('axios');
6

```

- Ab Zeile 26 definieren Sie die Funktion „WeatherHandler“, welche Sie vorher in Bild 19 angegeben haben. Zeile 26 gibt an, dass es sich dabei um diese Funktion handelt.

```

26 function weatherHandler(agent){
27   const apiKey = 'HIER API KEY EINSETZEN';
28   const stadt = agent.parameters['geo-city'];
29   return axios.get(`http://api.openweathermap.org/data/2.5/weather?q=${stadt}&appid=${apiKey}&units=metric`)
30   .then(result) => {
31     agent.add(`Die Temperatur in ${stadt} liegt bei ` + result.data.main.temp + ` Grad Celsius`);
32   };
33
34 }

```

- In Zeile 27 befindet sich der von uns eingefügte API-Schlüssel.
- In Zeile 28 wird die Variable „geo-city“ aus unserem „Wetter-Intent“ in die verständlichere Variable „stadt“ umgewandelt.
- In Zeile 29 fragt axios die API-Daten ab. Als erstes sehen Sie <http://api.openweathermap.org/data/> 2.5.weather gefolgt von einem „?“. Ab diesem Punkt haben Sie die Möglichkeit, der API unterschiedliche Parameter zuzuweisen.
- An erster Stelle befindet sich die Stadt, deren Wetterinformationen die Nutzer Ihres Agenten abfragen. Die Syntax der API schreibt vor, dass dafür q=Name der Stadt angeben wird. Um den Namen der abgefragten Stadt nicht wiederkehrend händisch eingeben zu müssen, nutzen Sie hier die Variable „stadt“, sodass Dialogflow automatisch die abgefragte Stadt einsetzt. Weitere Parameter werden mit einem „&“ voneinander getrennt. Zur Authentifizierung Ihrer Anfrage bei openweathermap.com benötigt Dialogflow dann Ihren API-Schlüssel. Um die Temperatur in Celsius anzeigen zu lassen, nutzen Sie im Code den Parameter units=metric.
- In Zeile 31 wandelt Dialogflow das für uns wichtigste Ergebnis, die Temp, um und gibt den angefragten Wert aus.

Happy coding!

Glossar

Action Eine Google Action ist eine Softwareapplikation für den Google Assistant. Ein strukturelles Analogon der Google Action ist der Skill für Amazons Alexa. Skills und Actions funktionieren ähnlich wie Apps für Mobile Devices. Die Auffindbarkeit ist allerdings derzeit in den Skill Stores noch unterentwickelt, Skills sind aktuell also noch schwerer auffindbar als Apps im Appstore.

Adobe XD Adobe XD ist ein Tool zur Konzeption von Interaktiven Anwendungen. Mit XD lassen sich komplexe Prototypen für Apps entwickeln. Gleichzeitig baut XD seine Möglichkeiten zur Entwicklung von Voice Prototypen aus.

Adversarial Attack Adversarial Attacks sind Inputs in ein Machine-Learning-Modell, die erstellt werden, um einen Fehler im System auszulösen. Sie funktionieren wie optische Täuschungen für Maschinen. Bei CUIs sind es akustische Informationen, die Angreifer einschleusen. Adversarial Attacks sind eines der Sicherheitsprobleme im Bereich AI. Problematisch ist vor allem, dass damit u. U. der Sprachsteuerung Befehle erteilt werden können, ohne dass der Nutzer es weiß.

Alexa Alexa ist der Name des von Amazon entwickelten Assistenten. Alexa ist auf den Amazon-eigenen Smartspeakern „Amazon Echo“, „Echo Show“ und „Echo Dot“ verfügbar. Alexa gibt es auch als Smartphone-App. Zudem wird Alexa auch an andere Unternehmen lizenziert und ist so beispielsweise in Autos, Smartspeakern und Mikrowellen-Geräten verfügbar.

Algorithmus Ein Algorithmus bezeichnet in der Informatik eine eindeutige Spezifikation, wie eine bestimmte Klasse von Problemen zu lösen ist. Algorithmen können beispielsweise Rechnungen ausführen, Daten verarbeiten oder automatisiert logische Argumente formulieren und logische Probleme lösen.

Angular.js Angular.js ist ein clientseitiges JavaScript-Web-Framework zur Erstellung von Single-Page-Webanwendungen nach einem Model-View-ViewModel-Muster. Es wird als OpenSource-Framework von Google bereitgestellt.

Application Programming Interface (API) Application Programming Interfaces, kurz: APIs, werden auf Deutsch oft als Programmierschnittstellen, gelegentlich auch

als Anwendungsschnittstellen bezeichnet. Korrekter wäre die Bezeichnung „Schnittstelle zur Programmierung von Anwendungen“. Diese Schnittstellen sind ein Programmteil, der es anderen Programmen ermöglicht, eine Anbindung an das Software-System herzustellen. Im Bereich von Onlinediensten ermöglicht die API, Webdienste zu nutzen und so Daten bzw. Content Dritter dynamisch für das eigene Programm zu verwenden.

Artificial Intelligence (AI, KI) Artificial Intelligence (AI), zu Deutsch: Künstliche Intelligenz, wird als Begriffskonstrukt verwendet, um die natürliche Intelligenz des Menschen von Lernprozessen einer Maschine zu unterscheiden. Der Begriff ist jedoch umstritten. Zum einen wird in Frage gestellt, inwieweit man die „Intelligenz“ der Maschine überhaupt mit der menschlichen Intelligenz, die ein Bewusstsein voraussetzt, vergleichen kann. Zum anderen gibt es die Überlegung, dass der Verweis auf „Künstliche Intelligenz“ lediglich ein kulturelles Konstrukt ist, das Dinge bezeichnet, die neu und wundersam erscheinen; die Zuschreibung würde fallengelassen, sobald sich die Technologie im kulturellen Mainstream etabliert habe (Tesler’s Theorem: „AI is whatever hasn’t been done yet.“). Derzeit werden als AI u. a. folgende Technologien bzw. technologische Forschungszweige klassifiziert: Spracherkennung und -verarbeitung, autonomes Fahren und so genannte „intelligente Agenten“, Geräte, die ihre Umwelt erfassen können und in (Re-)Aktionen mit den Ereignissen in ihrer Umwelt treten können, die dem Ziel bzw. der Erfüllung des Zwecks des Gerätes dienen.

Automatic Speech Recognition (ASR) Spracherkennung ist eine Teildisziplin der Computerlinguistik, die sich mit der Entwicklung von Methoden und Technologien befasst, die die Erkennung und Übersetzung gesprochener Sprache in Text durch Computer ermöglichen. Synonyme sind Automatic Speech Recognition (ASR) und Speech to Text (STT). Manche dieser Methoden und Technologien erfordern ein Training des Programms. Dabei analysiert das Programm die Stimme individueller Testpersonen und nutzt sie zur Optimierung der Spracherkennung. Big Data und Deep Learning haben zu einer erheblichen Beschleunigung der Entwicklung der Spracherkennung geführt, da nun extrem große Datenmengen zum Training und zur Auswertung von natürlicher Sprache herangezogen werden können.

Autoresponder Ein Autoresponder ist ein Programm, das vorgefertigte Antworten auf Anfragen automatisiert verschickt. Die ersten Chatbots waren Autoresponder, die eine Anfrage auf ein spezifisches Keyword untersuchten und ggf. eine spezifische vorgefertigte Antwort zurück gaben.

Bixby Bixby ist ein Virtueller Assistent, der von Samsung Electronics aus Südkorea entwickelt wurde. Bixby ist auf den Samsung-eigenen Smartphones, TV-Geräten und dem Smartspeaker („Bixby Home“) verfügbar.

Botsociety BotSociety ist ein Tool für die Konzeption von Chatbot-Mockups und Flussdiagrammen.

Brand Experience Brand Experience bezeichnet die Art und Weise, wie Nutzer eine Marke wahrnehmen und erleben. In Bezug auf Conversational User Interfaces stellt sich vor allem die Frage, wie diese klingen und sich ausdrücken müssen, um einer

Markenerfahrung zu entsprechen. Auch die Inhalte und die Funktionalität von Bots müssen die Werte einer Marke widerspiegeln.

Buzzword Buzzword ist der englische Ausdruck für ein Modewort, ein Begriff, der in aller Munde ist. Impliziert wird mit der Verwendung Buzzword auch, dass der Begriff bereits über Gebühr strapaziert wird.

Chatbot Chatbots sind eine Software, die in einem Conversational User Interface (beispielsweise einem Messenger-System wie Facebook Messenger oder dem Chinesischen WeChat) programmierte Nachrichten an User versenden.

Chatfuel Chatfuel ist ein Tool zur Erstellung von Chatbots.

Cognitive Services Cognitive Services sind technologische Plattformen, die eine natürliche Kommunikation mit Computern ermöglichen. Bestandteile dieser Plattformen sind unter anderem Machine Learning, NLP, Spracherkennung, Objekterkennung, Erkennung von Stimmungslagen von Menschen und Mensch-Maschine-Interaktion.

Cognitive Walkthrough Der Cognitive Walkthrough ist eine Methode aus der Usability-Evaluation. Sie basiert, wie der Name schon nahelegt, darauf, dass sich ein Usability-Experte in einen Nutzer hineinversetzt und das Interface aus dessen Sicht zu sehen versucht. Es wird hier auf ein empirisches Testing verzichtet. Der Cognitive Walkthrough besteht aus vier Schritten: 1. Definition des Inputs, 2. Untersuchung der Handlungssequenz, 3. Protokollieren kritischer Information, 4. Überarbeitung des Interfaces.

Confidence level Der Confidence Level gibt die prozentuale Wahrscheinlichkeit an, mit der eine Maschine bzw. ein Algorithmus ein Objekt „erkennt“. Dieses Erkennen funktioniert so, dass im Zuge von Machine Learning der Maschine Trainingsdaten zur Verfügung gestellt werden, in denen jeweilige Ziele markiert sind (Tagging). Die Maschine vergleicht diese Trainingsdaten mit aktuellen Daten, die ihr vorliegen und kommt zu einer Wahrscheinlichkeit, mit der sie beispielsweise ein Objekt auf einem Bild wiedererkennen. Dasselbe gilt auch für die Semantik eines Wortes oder die Bestimmung eines Intents. Diese Wahrscheinlichkeit beschreibt also die relative Gewissheit der Maschine.

Conversational Commerce Conversational Commerce ist ein Dialog zwischen einem Unternehmen und einem User mit Bezug auf Produkte oder Services und dem direkten oder indirekten Ziel, den User zum Kauf zu motivieren. In der Regel bezieht es sich auf die Konversation via digitalem Assistenten oder schriftbasiertem Chatbot. Anders als bei dem aktuell gängigen E-Commerce über Webshops auf dem Desktop oder dem mobilen Endgerät gibt es hier keine Ergebnislisten und auch keine umfangreichen Produktbeschreibungen. Dies schränkt die bereitgestellten Auswahlmöglichkeiten ein. Für den Nutzer kann dies entlastend wirken, da ein Zuviel an Auswahl die Nutzer eher überfordert (Paradox of Choice). Für Unternehmen bedeutet dies aber beispielsweise ein noch intensiveres Bemühen um Sichtbarkeit.

Conversational User Interfaces (CUI) Conversational User Interfaces (CUI) sind Interfaces, die Gespräche mit realen Menschen imitieren. CUIs sind nach den Graphic User Interfaces (GUI) die nächste große Disruption im Bereich der Interfaces. An die Stelle graphischer Interfaces, auf denen die User Eingaben vornehmen, tritt ein dialogischer, natürlichsprachiger Austausch mit dem User, der u. a. langwieriges Suchen in

hierarchisch strukturierten Websites obsolet macht. Für funktionierende CUIs sind Technologien wie Natural Language Processing und Natural Language Understanding entscheidend. Auch die Anbindung über APIs an Systeme Dritter ist ein wesentliches Erfolgskriterium.

Conversion Rate Die Conversion Rate ist eine KPI im Online-Marketing, die das Verhältnis der Anzahl der Besucher zur Anzahl der Conversions bezeichnet. Im E-Commerce ist eine Conversion in der Regel eine Kaufhandlung; was eine Conversion ist, kann aber für den jeweiligen Use Case definiert werden.

Cortana Cortana ist ein Virtueller Assistent, der von Microsoft entwickelt wurde. Cortana ist integraler Bestandteil der aktuellen Windows-Versionen und dadurch auf einer großen Zahl von Laptops verfügbar. Cortana ist ebenfalls Bestandteil der neueren Xbox-Geräte. Microsoft lizenziert Cortana auch an unabhängige Smartspeaker-Hersteller wie beispielsweise Harman Kardon.

Deep Learning (DL) Deep Learning ist eine Form der Machine-Learning-Algorithmen. Sie nutzen neuronale Netze mit einer Vielzahl an Ebenen zwischen Input und Output, in denen die empfangenen Daten weiterverarbeitet und zunehmend abstrakt werden. In diesen Ebenen laufen Prozesse des unbeaufsichtigten Lernens ab, etwa das Erkennen von Mustern in Eingabedaten, ohne dass ein Abgleich mit im Voraus bekannten Zieldaten stattfindet und ohne dass solche Daten bekannt sind.

Development-Plattform Eine Development-Plattform ist eine Entwicklungsumgebung, die von einem kommerziellen Anbieter zur Entwicklung von Virtuellen Assistenten, Chatbots, Skills und Actions zur Verfügung gestellt wird.

Dialogflow Dialogflow ist eine Plattform für die dialogorientierte Kommunikation und Natural Language Processing. Mit ihr lassen sich Chatbots realisieren und Anwendungen für den Google Assistant erstellen, die auch als Voice-Anwendungen funktionieren. Dialogflow ist in verschiedenen Editionen verfügbar und beherrscht zahlreiche Sprachen.

Duer Duer ist ein von der chinesischen Firma Baidu entwickelter virtueller Assistent, der Bestandteil der firmeneigenen Duer OS Betriebssystemumgebung ist. Baidu baut ein vollständiges Ökosystem mit Smartspeakern, Smart-Home-Komponenten und integriert dies über das Betriebssystem Duer, das über Voice gesteuert wird.

Engagement Engagement bezeichnet den Interaktionsgrad, der mit einem digitalen Inhalt erreicht wird. Engagement kann generell vieles sein – beispielsweise Klicks, Likes, Kommentare bzw. Interaktionen in einem Chatverlauf oder eine voicebasierte Konversation.

Entity Entities (Entitäten) sind Elemente des Natural Language Processing. Entitäten sind beispielsweise Personen, Organisationen, Orte und dergleichen.

Explizite Invocation Explicit Invocations (dt. explizite Invokationen) sind Invokationen (Aufrufe), bei denen der Nutzer einen Skill oder eine Action öffnet und eine konkrete Anwendung benennt, also z. B. „Alexa, öffne OTTO Unbox!“. Auch das Nutzen von Phrasen wie „Alexa....“ oder „OK Google“ kann als explizite Invokation begriffen werden. (Verwandt: implizite Invokation)

Findability Findability, zu Deutsch: Auffindbarkeit, bezeichnet den Grad an Einfachheit, mit der Inhalte durch User gefunden werden können. Für Sprachinterfaces ist Findability eine besondere Herausforderung, da Vorschläge relevanter Inhalte, die bei GUIs möglich und üblich sind, nicht in gleichem Maße funktionieren. Da Menschen nur Dinge fragen oder suchen können, von deren Existenz sie wissen, ergeben sich für Unternehmen u. a. neue Herausforderungen im Bereich SEO für Sprachinterfaces.

Flowchart Ein Flowchart ist ein Flussdiagramm. Flussdiagramme kommen beispielsweise bei der Gestaltung von Konversationen für Chatbots zum Einsatz.

Gartner Hype Cycle Der Hype Cycle bildet die Phasen der Aufmerksamkeit und Akzeptanz der Öffentlichkeit gegenüber einer neuen Technologie ab. Der Begriff wurde von der Gartner-Analystin Jackie Fenn geprägt. Auf der x-Achse wird die Zeit, auf der y-Achse der Grad an Aufmerksamkeit abgetragen. Der Beginn des Zyklus liegt im Nullpunkt (technischer Auslöser). Es folgt ein steiles Wachstum bis zum Maximum (Gipfel der überzogenen Erwartungen), anschließend ein ebenso steiler Abfall, allerdings nicht auf Null (Tal der Enttäuschungen). Anschließend steigt der Kurvenverlauf erneut (Pfad der Erleuchtung – realistischere Erwartungen an die technische Neuerung) und mündet schließlich ins Plateau der Produktivität, auf dem die Technologie solider wird und immer mehr allgemeine Akzeptanz erfährt.

Google Assistant Google Assistant ist ein von der Firma Google entwickelter virtueller Assistent, der eine Vielzahl der Google-Services (Karten, Suche, Kalender) integriert und sowohl auf den Google Android Phones, aber auch in Google Smartspeakern integriert ist. Google lizenziert den Assistent auch an Drittanbieter. Google Assistant lässt sich per Text oder Stimme steuern und zählt zu den leistungsfähigsten dieser Plattformen.

Hands-free „Hands-free“ ist ein Schlagwort, dass sich auf Situationen bezieht, in denen man die Hände für andere Dinge als zum Bedienen eines Graphic User Interfaces benötigt – beispielsweise beim Autofahren oder beim Kochen. Solche Hands-free Use Cases sind typisch für Voice User Interfaces.

HiAssistant HiAssistant ist ein von der chinesischen Firma Huawei entwickelter Virtueller Assistent.

Household Penetration (HHPEN) Household Penetration ist der Anteil aller Haushalte, die in einer zu definierenden geographischen Region von einem Service oder Produkt erreicht werden. „Erreicht“ bedeutet, der Dienst wird genutzt oder ein Gerät wird besessen.

Human Computer Interaction (HCI) Human-Computer-Interaction (HCI) oder auch Human-Machine-Interaction (HMI), auf Deutsch: Mensch-Maschine-Interaktion, befasst sich mit dem Design und der Nutzung von Interfaces, die die Schnittstelle zwischen Nutzern und Computern bilden. HCI ist als Disziplin an der Schnittstelle zwischen Informatik, Verhaltenswissenschaften, Design, Medienwissenschaften und diversen anderen Disziplinen angesiedelt.

Human-Centered Design Im Human-Centred Design geht es, wie der Name schon sagt, darum, die menschliche Perspektive in allen Schritten des Problemlösungsprozesses

ses mitzudenken. Human-Centered Design ist der Ausgangspunkt vieler Spielarten und Anwendungsfelder von Design. Im User-Interface- und User-Experience-Design ist ein Human-Centered-Ansatz besonders wichtig, weil Nutzer das Designprodukt ja nur erfahren können, wenn sie es auch nutzen können. Im Bereich von Conversational User Interfaces und vor allem Voice ist ein Human-Centered-Design-Ansatz besonders wichtig, weil das Interface minimalistisch bzw. „unsichtbar“ ist und der Nutzer daher wenig Möglichkeiten zur Orientierung und erklärenden Nutzung hat.

Implizite Invocation Eine implizite Invokation ist ein User Request, der nicht direkt eine Google Action mit dem Namen aufruft, sondern die einen Intent (eine Absicht) hat, die im Google Assistant auslöst, das eine passende Action geöffnet oder passende Inhalte ausgespielt werden. Beispiel: Was läuft im Fernsehen? Die Frage zielt auf keine konkrete Aktion. Der Assistent weiß jedoch, wo er die Informationen findet. In anderen virtuellen Assistenten gibt es ebenfalls implizite Invokationen.

Intelligent Assistant Ein intelligent Assistant (intelligenter Assistent) ist ein Software-Agent, der in der Lage ist, Aufgaben für einen Menschen zu erledigen (z. B. Bestellungen ausführen, Informationen beschaffen, Termine buchen).

Intent Ein Intent ist ein Element im Natural Language Processing. Es beschreibt die Absicht (Intention) die einer Nutzeräußerung zugrunde liegt.

Interactive Voice Response (System) (IVR) Ein Interactive Voice Response System (IVR) ermöglicht es den Usern, über akustische Medien natürlichsprachige Dialoge mit einer Maschine zu führen. Typische IVRs sind automatisierte Telefonhotlines, die Auswahlmöglichkeiten ansagen (z. B. „Sagen Sie: Rechnung, oder Buchung, oder legen Sie einfach auf“) und auf einfache Sprachbefehle reagieren. IVR-Systeme sind gewissermaßen rudimentäre Conversational User Interfaces, die bereits seit Ende der 80er-Jahre zunächst vor allem in den USA im Einsatz waren.

Internet of Things (IoT) Das Internet of Things (Deutsch: Internet der Dinge) bezeichnet eine Informationsinfrastruktur, die Gegenstände physischer oder virtueller Art über Sensoren miteinander vernetzt. Die Voice-Technologie spielt dabei eine zentrale Rolle, insofern die Schnittstelle zwischen Mensch und Informationsnetzwerk nicht selten über Sprache gelöst wird, beispielsweise bei Smart-Home-Anwendungen, wie Licht-, Klima- und Musiksteuerung.

Invocation (Word/Phrase) Eine Invocation (dt. Invokation) ist ein Wort oder eine Phrase mit der eine Google Action, ein Alexa Skill o. ä. geöffnet wird. Wird die Action oder der Skill direkt benannt und zur Öffnung aufgefordert, z. B. „Alexa, öffne OTTO Unbox!“), nennt man die Invokation direkt (explizit). Wird der Intent vom Assistenten auf eine Weise interpretiert, die ein Öffnen eines Skills oder einer Action zur Folge hat, nennt man die Invokation indirekt (implizit), z. B. „Alexa, ich möchte Kleider in meiner Nähe entdecken!“ – Alexa öffnet OTTO Unbox. Auf einer vorgesetzten Ebene ist aber auch das Aktivieren eines digitalen Assistenten über einen Sprachbefehl wie „Alexa....!“ oder „OK Google“ eine Invokation.

Key Performance Indicator (KPI) Ein Key Performance Indicator (KPI) ist eine qualifizierbare Größe, um den Erfolg einer Maßnahme, einer Dienstleistung, eines Produkts

oder einer Person zu messen. Ein KPI wird je nach messbarem Objekt individuell zusammengestellt und berechnet.

Kontext Der Begriff Kontext gehört ins Feld der Semantik. Je mehr Kontext Maschinen zur Verfügung haben, desto besser können diese beispielsweise User Intents interpretieren. Kontext macht Maschinen funktionaler. Beispiel: Wenn das Smartphone weiß, wo sich das Smartphone bzw. der Nutzer befindet, kann es relevantere Suchergebnisse, die im Umkreis sind, finden, als wenn es das nicht weiß. Dasselbe gilt für die Analyse von Utterances bei Natural Language Understanding und natürlich auch bei den Aktionen, die virtuelle Assistenten auslösen.

Labeling Labeling ist ein Grundinstrument maschinellen Lernens. Maschinen lernen durch das Markieren (Labeling) gewünschter Zieldaten, etwa das Markieren eines bestimmten Gegenstands in einem Bild. Die Maschine analysiert die Eigenarten des Bildes und vergleicht diese mit den Eigenarten eines anderen, möglicherweise ähnlichen Bildes. Dadurch lernt die Maschine durch jeden neuen Vergleich und kommt zu einem bestimmten Confidence Level, etwas erkannt zu haben. Beim Trainieren von Sprachdaten werden große Textmengen der Maschine als Trainingsdaten zur Verfügung gestellt, die diese mit der Sprache beispielsweise eines Spracheingabesystems vergleicht.

LUIS LUIS ist die Microsoft API für Natural Language Processing.

Machine Learning (ML) Machine Learning (ML) ist ein Teilstoff der Künstlichen Intelligenz, das statistische Methoden nutzt, um Computersystemen die Möglichkeit zu geben, etwas zu lernen. Lernen bezeichnet dabei die Fähigkeit, die Bewältigung einer gegebenen Aufgabe durch Verarbeitung neuer Daten zu verbessern, ohne explizit entsprechend programmiert worden zu sein.

Micro-App App (Applikation), die nur eine einzige Aufgabe erfüllt. Kann direkt auf dem Gerät ohne Installation aus einem App-Store gestartet werden.

Multimodal Multimodal bedeutet wörtlich „auf viele Weisen“. Im Kontext von CUI ist damit gemeint, dass Input und Output auf verschiedene Weisen, d. h. in verschiedenen Modi, gestaltet sein können, je nach Anwendungskontext. Zum Einsatz können dabei Voice-Elemente kommen, aber natürlich auch graphische Interfaces (Bildschirme) oder Sensoren, wenn man an den Bereich IoT denkt.

Natural Language Generation (NLG) Natural Language Generation (dt. Textgenerierung, auch: natürlich-sprachliche Generierung) meint die automatische Generierung natürlicher Sprache durch eine Maschine. NLG ist Teil der Computerlinguistik und wird häufig unter dem Oberbegriff „künstliche Intelligenz“ subsumiert.

Natural Language Processing (NLP) Natural Language Processing (NLP) befasst sich mit der Verarbeitung von natürlicher Sprache in der Mensch-Maschine-Interaktion. Schwerpunkt der Forschung ist die Frage, wie Computer so programmiert werden können, dass sie in der Lage sind, natürlichsprachige Daten in großen Mengen zu verarbeiten und zu analysieren. NLP beinhaltet Problemfelder wie Spracherkennung, Natural Language Understanding und Natural Language Generation.

Natural Language Understanding (NLU) Natural Language Understanding ist ein Unterthema vom Natural Language Processing (NLU). Im NLU geht es um die Verarbeitung und Interpretation natürlicher Sprache durch Maschinen. Dabei kann die Eingabe als schriftsprachliche oder natürlichsprachliche Äußerung erfolgen.

Natural User Interfaces Natural User Interfaces (NUI) ermöglichen dem Nutzer eine Befehlseingabe ohne Hilfsmittel. Die Befehle werden haptisch (Wischen mit dem Finger/Tippen), akustisch (Spracherkennung) oder optisch (Gesichtserkennung) von der Maschine aufgenommen bzw. erkannt.

Neural Networks (NN, artificial neural network (ANN) or simulated neural network (SNN)) Neural Networks (dt. Neuronale Netzwerke) sind nicht lineare, statistische Datenmodellierungs- oder Entscheidungsfindungsinstrumente. Sie werden genutzt, um komplexe Beziehungen zwischen Inputs und Outputs zu modellieren oder Muster in Daten zu erkennen. Neuronale Netze bestehen aus einer Vielzahl an Ebenen zwischen Input und Output, in denen die empfangenen Daten weiterverarbeitet werden. In diesen Ebenen laufen Prozesse des unbeaufsichtigten Lernens ab, etwa das Erkennen von Mustern in Eingabedaten, ohne dass ein Abgleich mit im Voraus bekannten Zieldaten stattfindet und ohne dass solche Daten bekannt sind.

node.js Node.js ist eine Plattform in der Softwareentwicklung zum Betrieb von Netzwerkanwendungen, mit der v. a. Webserver realisiert werden können. Node.js kommt häufig zum Einsatz, wenn man komplexe Chatbots entwickelt und dabei beispielsweise die NLP-Engine eines Anbieters mit den eigenen Daten und der Schnittstelle einer Distributionsplattform verbindet. Mit Node.js kann man beispielsweise einen Facebook Chatbot mit dem Google Dialog Flow-NLP betreiben.

On Device/On the cloud „On Device“ vs. „on the cloud“ bezieht sich auf den Ort der Datenverarbeitung, wo also die Sprach- oder Bilderkennung stattfindet. Auf dem Gerät („on device“) können diese Prozesse bis zu zehn Mal schneller erfolgen als „on the cloud“, was die Dialoge zwischen Mensch und Maschine dramatisch beschleunigt und damit verbessert. So wie die Google-Suche nicht zuletzt durch ihre Performance überzeugt und zum Suchen anregt, ist auch die Reaktionsgeschwindigkeit („Latency“) bei der Spracherkennung ein entscheidender Faktor. Der Google Assistant kann Automatic Speech Recognition auf modernen Android-Phones realisieren und lässt den Nutzer damit das gesamte Telefon durch Sprache steuern, was die Steuerung durch Touch bzw. Click veraltet und langsam erscheinen lässt. On-Device-Verarbeitung kann auch in verschiedenen Smart-Home-Geräten und letztlich auch im Smartspeaker erfolgen, immer mit dem Vorteil der Beschleunigung der Interaktion.

Personal Digital Assistant (PDA) Ein Personal Digital Assistant (Persönlicher Digitaler Assistent) ist ein Software-Agent, der in der Lage ist, Aufgaben für einen Menschen zu erledigen (z. B. Bestellungen ausführen, Informationen beschaffen, Termine buchen).

Personal Navigational Device (PND) Ein Personal Navigational Device ist ein elektronisches Gerät, das eine Erfassung von Positionsdaten, z. B. via GPS, mit einer Navigationsoption verbindet. Er kann als reduzierte Form des Personal Digital Assistants begriffen werden.

Prototype Ein Prototype (dt. Prototyp) ist ein funktionsfähiges Versuchsmodell eines geplanten Produktes. Prototypen für Voice- bzw. Conversational-Anwendungen unterscheiden sich von anderen Prototypen – beispielsweise für Smartphone Apps – stark, da die Pfade nicht immer gleich sind. Hier kommen verschiedene Prototyping-Tools wie Botsociety, Adobe XD, oder auch BotMock zum Einsatz. In frühen Entwicklungsstadien von Prototypes eignen sich auch Verfahrensweisen wie Wizard of Oz um für Entwickler und anderer Stakeholder einer Conversational Application einen ersten Eindruck über die Funktionsweise zu erhalten.

Prototyping Prototyping ist der Prozess der Entwicklung ein funktionsfähiges Versuchsmodell eines geplanten Produktes. Eine erste Version wird auch MVP (Minimal Viable Product/minimal überlebensfähiges Produkt) genannt. Es ist die erste funktionsfähige Version des Produkts oder Services.

Rasa Rasa ist eine deutsche Open-Source-Entwicklungsumgebung für Chatbots und virtuelle Assistenten, die international eingesetzt wird. Mit Rasa können Entwickler eigene NLP-Engines aufbauen, trainieren und betreiben und damit den Einsatz von Anwendungen der großen Unternehmen (Facebook, Google, Amazon etc.) vermeiden.

react.js React ist eine von Facebook entwickelte Softwarebibliothek in JavaScript. Sie bietet eine Basis für die Ausgabe von User-Interface-Komponenten von Websites.

Semantisch Semantisch ist ein linguistischer Begriff und heißt „die Bedeutung der Zeichen (Worte) betreffend“. Das dazugehörige Nomen ist Semantik (Bedeutungslehre). In unserem Kontext bezieht sich „semantisch“ meist auf das „Semantic Web“, bei dem es im Wesentlichen um die Ergänzung von vorhandenen Inhalten durch strukturierte Daten geht. Strukturierte Daten spezifizieren im Code die Bedeutung eines Begriffs und tragen so dazu bei, dass Maschinen den Kontext von Begriffen besser verstehen können. Semantik im informatischen Sinne ist also grundlegend für die Weiterentwicklung von NLU, NLP und NLG. Für Conversational User Interfaces ist dies ein besonders relevanter Aspekt, da Menschen bei der Form der Mensch-Maschine-Interaktion mit Maschinen annähernd so agieren wie mit realen Menschen. Entsprechend verwenden sie eher eine natürliche Sprache, die von Maschinen verstanden werden muss. Auch die Reaktion muss entsprechend natürlchsprachig wirken, um die User-Erwartungen nicht zu enttäuschen.

Sentiment Analyse Man könnte den Begriff mit Stimmungs- oder Gefühlsanalyse übersetzen. Es geht um die Deutung einer Aussage in Hinsicht auf ihre Polarität. Beschreibt ein Text das Objekt positiv, negativ oder neutral? Zur Einordnung werden Begriffe herangezogen wie „super“, „genial“, „ganz mies“ oder „totaler Schrott“. Eine zu ermittelnde Abstufung ist nicht immer eindeutig, so wird beispielsweise das Wort „perfekt“ von Menschen im Schnitt positiver wahrgenommen als das Wort „zufrieden“, obwohl beide im positiven Bereich liegen. Für die maschinelle Analyse einer Aussage sind Ironie und Zynismus noch eine Hürde.

Siri Siri ist ein von der Firma Apple entwickelter virtueller Assistent, der durch IOS allen modernen Apple-Geräten zur Verfügung steht.

Skill Eine Alexa Skill ist eine Softwareapplikation für Amazons intelligenten Assistenten Alexa. Ein strukturelles Analogon des Alexa Skills ist die Google Action. Skills und Actions funktionieren ähnlich wie Apps für Mobile Devices. Die Auffindbarkeit ist allerdings derzeit in den Skill-Stores noch unterentwickelt. Skills sind aktuell noch schwerer auffindbar als Apps im Appstore.

Smart Display Ein Smart Display ist ein Smartspeaker mit einem kleinen Bildschirm. Ein Beispiel für ein Smart Display ist der Amazon Echo Show. Die Mitbewerber von Amazon haben analoge Produkte auf dem Markt oder in Planung.

Smartspeaker Ein Smartspeaker ist ein kleiner Tischlautsprecher und ein Voice-Command-Gerät mit integriertem intelligenten Assistenten. Dies ermöglicht eine Aktivierung des Assistenten und zugehöriger Anwendungen mit Hilfe von Invocations. Die Ausstattung und der Funktionsumfang variieren von Hersteller zu Hersteller.

Speech Recognition Die Spracherkennung ist ein Teilgebiet der angewandten Informatik, der Ingenieurwissenschaften und der Computerlinguistik. Sie beschäftigt sich mit der Untersuchung und Entwicklung von Verfahren, die Automaten, insbesondere Computern, die gesprochene Sprache der automatischen Datenerfassung zugänglich macht. Die Spracherkennung ist zu unterscheiden von der Stimm- bzw. Sprecherkennung, einem biometrischen Verfahren zur Personenidentifikation. Allerdings ähneln sich die Realisierungen dieser Verfahren.

Speech to Text (STT) Speech to Text, auch: Voice to Text, beschreibt den Prozess der Konvertierung von natürlichsprachigem Input in digitalen Text mittels Spracherkennung.

Supervised learning Überwachtes Lernen gehört zum Bereich „maschinelles Lernen“ und ist der Komplementärbegriff zu unüberwachtem Lernen (engl. „unsupervised learning“). Zum überwachten Lernen gehört die Fähigkeit einer KI, Hypothesen über Gesetzmäßigkeiten zu bilden. Überwacht ist dieses Lernen, weil die Maschine anhand von Naturgesetzen oder Expertenwissen trainiert wird. Dabei sind die korrekten Ergebnisse also bekannt und der Lernprozess der Maschine kann anhand der bekannten Ergebnisse überwacht werden.

Text to Speech (TTS) Text to Speech ist eine Form der Sprachsynthese, bei der schriftlicher Text in gesprochenen Voice-Output konvertiert wird.

TingTing TingTing ist ein Smartspeaker des chinesischen Herstellers Tencent mit Integration von WeChat, das ebenfalls Tencent gehört.

TMall Genie TMall Genie ist ein virtueller Assistent, der von dem chinesischen Hersteller Alibaba entwickelt wurde.

Tonalität Tonalität bezeichnet die Art und Weise des sprachlichen Ausdrucks, der Darstellung, die beeinflusst, wie ein Sachinhalt wahrgenommen wird. Im Bereich Conversational User Interfaces ist die Tonalität ein wichtiger Faktor, da visuelle Elemente zugunsten des natürlichsprachigen Ausdrucks in den Hintergrund treten. Gerade für das Branding von Unternehmen ist die Frage der Tonalität bei digitalen Assistenten essentiell.

Training (Trainingsdaten) Trainingsdaten sind in unserem Fall Daten, die einer KI vorlegt werden, um schrittweise das Confidence Level der KI zu erhöhen. Je mehr ge-

labelte Daten die KI zum Abgleich zur Verfügung hat, desto sicherer wird sie in der Identifikation.

Turing-Test Der Turing-Test ist nach Alan Turing benannt, der mit diesem Test 1950 ermitteln wollte, ob Computer ein dem menschlichen Denken gleichwertiges Denkvermögen haben. Für den Test muss ein menschlicher Fragesteller über eine Tastatur eine Unterhaltung mit zwei Unbekannten führen. Er hat dabei keinen Sicht- und keinen Hörkontakt. Was der Proband während des Gesprächs nicht weiß ist, dass einer der beiden Gesprächspartner ein Computer ist. Kann der Proband am Ende nicht eindeutig benennen, welche der beiden Gesprächspartner der Computer, welcher der Mensch war, hat der Computer den Turing-Test bestanden und es wird ihm ein dem Menschen analoges Denkvermögen attribuiert.

Uncanny Valley In der Ästhetik bezeichnet das „Uncanny Valley“ eine Hypothese über das Verhältnis zwischen dem Grad an Ähnlichkeit eines Objekts mit einem Menschen und der emotionalen Reaktion auf dieses Objekt. Während normalerweise die Affinität mit steigender Ähnlichkeit zunimmt, lösen Objekte, die fast täuschend menschliche Züge haben, eher Ablehnung und ein Gefühl des Unheimlichen aus.

Unsupervised learning Unüberwachtes Lernen gehört zum Bereich maschinelles Lernen und ist der Komplementärbegriff zu überwachtem Lernen (engl. „supervised learning“). Zum unüberwachten Lernen gehört Erkennen von Mustern in Eingabedaten, ohne dass ein Abgleich mit im Voraus bekannten Zieldaten stattfindet und ohne dass solche Daten bekannt sind.

Usability Usability bedeutet auf Deutsch „Gebrauchstauglichkeit“. Damit wird das Ausmaß bezeichnet, in dem ein Produkt, Dienst oder System in einem klar definierten Anwendungskontext effektiv zur Erreichung seiner Ziele eingesetzt werden kann. Usability ist eng verwandt mit UX.

Usability Testing Usability Tests sind empirische Verfahren, um die Gebrauchsfertigkeit von Soft- oder Hardware zu überprüfen. Ihr Gegenpart ist der Cognitive Walkthrough, der ohne empirische Tests mit echten Usern auskommt.

User Experience (UX) User Experience (UX) wird im Deutschen als Nutzererlebnis, Nutzungserlebnis oder Anwendererlebnis bezeichnet. UX umfasst alle Aspekte der Erfahrung, die ein User in der Interaktion mit einem Produkt, einem Dienst, einer Umgebung oder einer Einrichtung macht. Software und IT-Systeme zählen ebenfalls zu diesen Interaktionspunkten und bezieht sich oft auf die Gestaltung von Websites oder Apps, kann aber auch die nicht-digitale, physische Nutzung mit einschließen.

User Interface (UI) Das User Interface (dt. „Benutzerschnittstelle“, „Nutzerschnittstelle“ oder auch „Benutzungsschnittstelle“) ist jeder Ort oder jede Handlung, an dem oder mit der ein User mit einer Maschine in Interaktion tritt.

User Journey Die User Journey ist ein Begriff aus dem Feld der User Experience. Er beschreibt die Phasen und Schritte, die der Nutzer einer interaktiven Anwendung durchläuft, bis er am Ziel seines Nutzungsvorgangs angekommen ist, oder die Anwendung aus anderen Gründen verlässt. Häufig werden in der User Journey auch Schritte vor der Nutzung der Anwendung und Schritte danach erfasst und beschrieben. Die

Untersuchung und Optimierung der User Journey gehört zu den wichtigen Aufgaben bei Planung, Umsetzung und Monitoring von Anwendungen. Bei Conversational User Interfaces sind die User Journeys häufig schwerer zu beobachten, als bei Apps und Websites, da Conversational User Interfaces im besten Fall 100 Prozent personalisierte Dialoge mit dem Nutzer führen.

Utterance Utterance bedeutet auf Deutsch „Äußerung“. In unserem Fall geht es um die Äußerung eines Nutzers. Diese kann ein Wunsch, ein Befehl, eine Frage und vieles mehr sein. Eine Utterance kann isoliert stehen oder im Kontext eines vorangegangenen Wunsches, Frage, Befehl etc. Es ist für Maschinen schwer, Utterances einzuordnen. Schon allein die Entscheidung, wann eine Utterance beginnt und wann diese endet, ist komplex. Redepausen oder Unterbrechungen können den Eindruck erwecken, eine Utterance sei beendet. Andererseits wäre es wünschenswert, wenn Maschinen mit der Reaktion auf eine Utterance schon beginnen, ehe diese ganz zuende ist. Menschen tun dies ganz selbstverständlich, da bereits während der Gegenüber einen Satz spricht mit der Interpretation und dem Verständnis seiner Aussage begonnen wird. Unter Umständen wird diese Äußerung unseres Gegenübers sogar durch bestätigendes „Ja!“, „Hm-mm...“ oder „Genau!“ unterstützt bzw. unterbrochen. Solche Ansätze gibt es auch bereits im NLP.

Virtual Assistant Ein Virtual Assistant (Virtueller Assistent) ist ein Software-Agent, der in der Lage ist, Aufgaben für einen Menschen zu erledigen (z. B. Bestellungen ausführen, Informationen beschaffen, Termine buchen). Manchmal wird das Wort „Chatbot“ als Synonym für „Virtual Assistant“ verwendet; häufiger meint „Chatbot“ jedoch solche virtuellen Assistenten, die über einen schriftsprachigen Online-Chat zugänglich sind.

Voice Assistant Ein Voice Assistant ist ein Conversational User Interface, dem der Nutzer in natürlicher Sprache Anweisungen geben und Wünsche mitteilen kann, die der Assistent umsetzt. So kann ein Voice Assistant beispielsweise auf Sprachbefehl eine Zugverbindung heraussuchen und ein passendes Ticket kaufen.

Voice Commerce Voice Commerce bezeichnet E-Commerce durch Sprachbefehle. Voice Commerce unterscheidet sich von E-Commerce dadurch, dass bei Sprachbefehlen die Komplexität der Antwort des Voice-Commerce-Systems naturgemäß reduziert sein muss, da kein Bildschirm zur Verfügung steht und damit Auswahlmöglichkeiten stark reduziert werden müssen. Fragen der Authentifizierung beim Kaufvorgang stellen gegenwärtig noch Herausforderungen für Voice Commerce dar.

Voice Prototype Ein Voice Prototype ist ein funktionsfähiges Versuchsmodell eines geplanten Voice-Produktes. Prototypen für Voice-Anwendungen unterscheiden sich von anderen Prototypen – beispielsweise für Smartphone Apps – stark, da die Pfade nicht immer gleich sind. Hier kommen verschiedene Prototyping-Tools wie Botsociety, Adobe XD oder auch BotMock zum Einsatz. In frühen Entwicklungsstadien von Prototypen eignen sich auch Verfahrensweisen wie Wizard of Oz um für Entwickler und andere Stakeholder einer Conversational Application einen ersten Eindruck über die Funktionsweise zu erhalten.

Voice Search Voice Search bedeutet, dass der Nutzer seine Stimme (und natürliche Sprache) – anstelle von Touchscreen oder Tastatur – nutzt, um eine Suchanfrage im Internet, auf einer Website oder in einer App zu stellen. Die Recherche über intelligente Assistenten ist ein Anwendungsszenario für Voice Search. Voice Search ist oft interaktiv und umfasst mehrere Interaktionsrunden, ist also dialogisch.

Voice SEO Voice Search Optimization bezeichnet die Optimierung von Inhalten für die Sprachsuche. Voice SEO ist damit die konsequente Weiterführung von Suchmaschinenoptimierung in einer Zeit, in der die Sprachsuche an Bedeutung gewinnt. In der Vergangenheit wurden Suchanfragen üblicherweise auf der Tastatur oder dem Touchscreen eingegeben. Sie umfassen daher in der Regel nur wenige Schlüsselworte. Dies ändert sich bei der natürlichsprachigen Spracheingabe. Voice SEO bedeutet darum insbesondere das Optimieren auf natürlichsprachige Queries der User. Fragen und langen Phrasen kommt entsprechend eine größere Bedeutung zu.

Voice to Text Voice to Text, auch: Speech to Text, meint den Prozess der Konvertierung von natürlichsprachigem Input in digitalen Text mittels Spracherkennung.

Voice User Experience (VUX) Voice User Experience beinhaltet alle Aspekte der Erfahrungen, die ein Nutzer bei der Interaktion mit einem Produkt, Dienst, seiner Umgebung oder Einrichtung macht, die Voice-Technologie nutzt.

Voice User Interface (VUI) Voice User Interfaces (VUI) ermöglichen die gesprochene Interaktion zwischen Mensch und Computer, indem Spracherkennung gesprochene Befehle erkennt und üblicherweise Text-to-Speech-Technologie nutzt, um die Antwort auszuspielen. Ein Gerät, das mit einem VUI betrieben wird, bezeichnet man auch als Voice Command Device (VCD). VUI findet derzeit bereits Anwendung in Fahrzeugen, Haushaltsgeräten und vielem anderem mehr.

Watson Watson ist ein von IBM entwickeltes Computerprogramm, das unter anderem Antworten auf Fragen in natürlicher Sprache ausgeben kann. IBM Watson gewann 2011 eine Partie des Fernsehquizzes „Jeopardy!“ gegen zwei menschliche Gegner und brachte das Thema Artificial Intelligence ins Bewusstsein einer breiteren Öffentlichkeit. Unter der Marke Watson bündelt IBM eine ganze Reihe kognitiver Services und bietet Branchenlösungen – beispielsweise in der Medizintechnik – an.

Watson Assistant Watson Assistant ist ein Tool, mit dem Conversational User Interfaces für jede Art von App, Gerät oder Kanal entwickelt werden können.

WeChat WeChat ist eine weit verbreitete chinesische Messenger-Plattform, die weit hin als Rollenmodell angesehen wird. Während die Grundfunktionen von WeChat vergleichbar mit dem in Europa verbreiteten WhatsApp sind, gehen die Funktionalitäten von WeChat weit über das Austauschen persönlicher Nachrichten hinaus. Auf WeChat kann man auch mit Firmen und Institutionen „chatten“. Dabei bieten diese Kommunikationspartner Funktionen im Chatverlauf an, die stark an Funktionen aus Apps erinnern. Allerdings mit dem Unterschied, dass Nutzer den Messenger-Dienst nicht verlassen müssen und eine App downloaden, suchen usf. müssen, sondern die Funktionen direkt im Dialog mit dem Anbieter zur Verfügung gestellt bekommen. So kann man eine Institution im Messenger fragen, wo ihr Büro ist und die Antwort ist

eine funktionale Karte im Chatverlauf, wie man es von Google Maps gewohnt ist. Auf diese Weise entsteht ein Nutzerinterface für mobile Endgeräte, das sich stark von dem im gewöhnlichen App-Interface westlicher Prägung unterscheidet. Unterstützt werden diese Funktionen durch die Bezahlfunktion, die zu einem Kernfeature von WeChat gehört und mit der man in China praktisch überall, sogar beim Kauf einer Suppe am Straßenrand, bargeldlos bezahlen kann. WeChat war Vorreiter bei der Einführung von Chatbots und bietet heute umfangreiche Sprachsteuerungsfunktionen an.

Wizard of Oz Das Wizard-of-Oz-Experiment ist ein Instrument zum sehr frühen Prototyping von Sprachanwendungen. Dabei führt ein Proband vorgeblich ein Gespräch mit einer Maschine, während tatsächlich ein echter Interaktionspartner auf der anderen Seite sitzt, der das Verhalten des Probanden, seine Reaktionen auf die Sprachanwendung, analysiert und späterhin auswertet.

Wolfram Alpha Wolfram Alpha ist ein Onlinedienst zum Auffinden und Verarbeiten von Informationen. Anders als bei normalen Suchmaschinen werden keine Ergebnislisten mit Inhalten im Web erstellt, sondern stattdessen Daten durch den Algorithmus verarbeitet und dem Nutzer aufbereitet präsentiert. Natural Language Processing sowie Semantik sind der Schlüssel zur Funktion dieses Onlinedienstes.

Xiao Xiao ist ein virtueller Assistent, der von dem chinesischen Hersteller Xiaomi entwickelt wurde und in dessen Smartphones installiert ist. Xiaomi bietet auch eine Reihe von Smartspeakern und Smart-Home-Geräten an, in denen Xiao verfügbar ist.

Xiaowei Xiaowei ist ein virtueller Assistent, der von der chinesischen Firma Tencent entwickelt wurde.