

Tas, Serpil; Hildebrandt, Christian; Arnold, René

**Working Paper**

**Sprachassistenten in Deutschland**

WIK Diskussionsbeitrag, No. 441

**Provided in Cooperation with:**

WIK Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste GmbH, Bad Honnef

*Suggested Citation:* Tas, Serpil; Hildebrandt, Christian; Arnold, René (2019) : Sprachassistenten in Deutschland, WIK Diskussionsbeitrag, No. 441, WIK Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste, Bad Honnef

This Version is available at:

<https://hdl.handle.net/10419/227052>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Sprachassistenten in Deutschland

Autoren:  
Serpil Taş  
Christian Hildebrandt  
René Arnold

Bad Honnef, Juli 2019

## Impressum

WIK Wissenschaftliches Institut für  
Infrastruktur und Kommunikationsdienste GmbH  
Rhöndorfer Str. 68  
53604 Bad Honnef  
Deutschland  
Tel.: +49 2224 9225-0  
Fax: +49 2224 9225-63  
E-Mail: [info@wik.org](mailto:info@wik.org)  
[www.wik.org](http://www.wik.org)

### Vertretungs- und zeichnungsberechtigte Personen

Geschäftsführerin und Direktorin	Dr. Cara Schwarz-Schilling
Direktor Abteilungsleiter Post und Logistik	Alex Kalevi Dieke
Direktor Abteilungsleiter Netze und Kosten	Dr. Thomas Plückebaum
Direktor Abteilungsleiter Regulierung und Wettbewerb	Dr. Bernd Sörries
Leiter der Verwaltung	Karl-Hubert Strüver
Vorsitzende des Aufsichtsrates	Dr. Daniela Brönstrup
Handelsregister	Amtsgericht Siegburg, HRB 7225
Steuer-Nr.	222/5751/0722
Umsatzsteueridentifikations-Nr.	DE 123 383 795

In den vom WIK herausgegebenen Diskussionsbeiträgen erscheinen in loser Folge Aufsätze und Vorträge von Mitarbeitern des Instituts sowie ausgewählte Zwischen- und Abschlussberichte von durchgeführten Forschungsprojekten. Mit der Herausgabe dieser Reihe bezweckt das WIK, über seine Tätigkeit zu informieren, Diskussionsanstöße zu geben, aber auch Anregungen von außen zu empfangen. Kritik und Kommentare sind deshalb jederzeit willkommen. Die in den verschiedenen Beiträgen zum Ausdruck kommenden Ansichten geben ausschließlich die Meinung der jeweiligen Autoren wieder. WIK behält sich alle Rechte vor. Ohne ausdrückliche schriftliche Genehmigung des WIK ist es auch nicht gestattet, das Werk oder Teile daraus in irgendeiner Form (Fotokopie, Mikrofilm oder einem anderen Verfahren) zu vervielfältigen oder unter Verwendung elektronischer Systeme zu verarbeiten oder zu verbreiten.

ISSN 1865-8997

**Inhaltsverzeichnis**

<b>Abbildungsverzeichnis</b>	<b>IV</b>
<b>Tabellenverzeichnis</b>	<b>V</b>
<b>Abkürzungsverzeichnis</b>	<b>VI</b>
<b>Zusammenfassung</b>	<b>VII</b>
<b>Summary</b>	<b>VIII</b>
<b>1 Einleitung</b>	<b>1</b>
<b>2 Überblick zu Sprachschnittstellen und -assistenten</b>	<b>2</b>
2.1 Definition und Abgrenzung verschiedener Typen von Sprachschnittstellen und -assistenten	2
2.2 Momentaufnahme der Funktionen und Eingriffsmöglichkeiten für den Nutzer	10
2.3 Sprachassistenten und Verbraucherschutz	14
2.3.1 Datenschutz	14
2.3.2 Sicherheitsaspekte von Sprachassistenten	16
<b>3 Nutzungsintensität und Nutzungsmuster von Sprachassistenten in Deutschland</b>	<b>19</b>
3.1 Marktdurchdringung und Marktanteile	19
3.2 Nutzungsmuster	26
3.2.1 Funktionen	26
3.2.2 Multi-homing	28
3.2.3 Konfiguration	32
3.3 Beziehung zu anderen Diensten und Endgeräten	35
<b>4 Einstellung zu Sprachassistenten</b>	<b>38</b>
<b>5 Herausforderung für Wettbewerb und Regulierung</b>	<b>48</b>
<b>6 Fazit</b>	<b>53</b>
<b>Literaturverzeichnis</b>	<b>54</b>
<b>Anhang: Methodik</b>	<b>60</b>

## Abbildungsverzeichnis

Abbildung 2-1:	Beispiele für verschiedene populäre Sprachassistenten aus Patentanmeldungen der Hersteller	5
Abbildung 2-2:	Das Amazon Alexa-Ökosystem	7
Abbildung 3-1:	Zeitliche Entwicklung von intelligenten Sprachschnittstellen	20
Abbildung 3-2:	Anteil der Internetnutzer, die einen Sprachassistenten genutzt haben – Europa (Q4 2018)	21
Abbildung 3-3:	Rogers Innovationsadoptionskurve – Sprachassistenten*	22
Abbildung 3-4:	Adoption von Sprachassistenten	24
Abbildung 3-5:	Nutzungshäufigkeit: Sprachassistenten*	25
Abbildung 3-6:	Genutzte Funktionen – Sprachassistenten*	26
Abbildung 3-7:	Genutzte Funktion nach Sprachassistenten	27
Abbildung 3-8:	Muti-homing – Sprachassistenten*	29
Abbildung 3-9:	Multi-homing: Funktionen	30
Abbildung 3-10:	Multi-homing: Funktionen – Alexa und Google Assistant	31
Abbildung 3-11:	Multi-homing: Funktionen – Alexa und Siri	31
Abbildung 3-12:	Multi-homing: Funktionen – Google Assistant und Cortana	32
Abbildung 3-13:	Änderung von Einstellungen	33
Abbildung 3-14:	Gründe, warum keine (weiteren) Einstellungen der Sprachassistenten* geändert wurden	34
Abbildung 3-15:	Einkauf über Alexa	36
Abbildung 4-1:	Nutzung der einzelnen Dienste und Gesamt	38
Abbildung 4-2:	Einstellung zu Sprachassistenten*: Allgemein – Kenner	39
Abbildung 4-3:	Einstellung zu Sprachassistenten – Kenner	40
Abbildung 4-4:	Einstellung zu Sprachassistenten*: Allgemein – Alle	42
Abbildung 4-5:	Gründe für das Aufhören der Nutzung – ehemalige Nutzer	43
Abbildung 4-6:	Zukünftige Nutzungsintention: Kenner aber Nicht-Nutzer (Noch Nie)	47
Abbildung 5-1:	Netzwerk – und Feedbackeffekte von Sprachassistenten	49
Abbildung 5-2:	Vergleich der Spracherkennungsfehlerrate mit der Anzahl der vorhandenen Sprachinteraktionen mit Nutzern	50

## Tabellenverzeichnis

Tabelle 2-1:	Überblick zur Definition von Sprachschnittstellen bzw. Sprachdialogsystemen verschiedener Ausprägung in der Literatur	4
Tabelle 2-2:	Ergebnisse der Clusteranalyse von Knotte et al. (2019)	8
Tabelle 2-3:	Übersicht zu dokumentierten erfolgreichen Angriffen auf Sprachassistenten über die Sprachschnittstelle	17
Tabelle 3-1:	Kombination von Sprachassistenten	29
Tabelle 3-2:	Änderung von Einstellungen: Stimme	33

## Abkürzungsverzeichnis

CRF	Conditional Random Field
DNN	Deep Neural Networks
GUI	Graphical User Interface
IVR	Interactive Voice Response
HMM	Hidden Markov Model
NLP	Natural Language Processing
RNN	Recurrent Neural Network
SBS	Stimmliche Benutzerschnittstelle
SDS	Spoken Dialogue System
VUI	Voice User Interface
WWW	World Wide Web

## Zusammenfassung

Ob Amazons Alexa, Apples Siri, Googles Assistant, Microsofts Cortana oder Samsungs Bixby: Sprachassistenten werden immer populärer. Laut der Befragung von 3.184 Konsumenten in Deutschland, die für diesen Diskussionsbeitrag durchgeführt wurde, nutzten Ende des Jahres 2018 schon 26% zumindest einen solchen Sprachassistenten. Die Adoptionsrate ähnelt damit derjenigen von Smartphones etwa fünf Jahre nach Einführung der wesentlichen ersten Modelle.

Der entscheidende Unterschied zwischen den beiden Technologien ist, dass die meisten Konsumenten in Deutschland keinen Cent zusätzlich ausgeben und nicht einmal eine zusätzliche App herunterladen müssen, um einen Sprachassistenten zu nutzen. Rund 85% haben schon zumindest ein Endgerät zuhause, das einen Sprachassistenten vorinstalliert hat. Oft handelt es sich dabei um das Smartphone oder den Laptop. Bei etwa 11% der Befragten sind es spezielle sogenannte Smart Speaker.

Mit dieser großen Verbreitung von Endgeräten mit vorinstallierten integrierten Sprachassistenten ist ein sprunghafter Anstieg der Adoptionsrate möglich. Wahrscheinlich ist ein solcher Anstieg nach den Ergebnissen der Befragung in naher Zukunft jedoch nicht. Die überwiegende Mehrheit der Nicht-Nutzer hat nicht die Absicht, im nächsten Jahr einen der fünf populärsten Sprachassistenten zu verwenden. Die derzeitigen Nutzungsmuster deuten ebenso nicht auf einen großen Einfluss der Assistenten hin. Zumeist werden sie eher selten und für eine recht eingeschränkte Auswahl von ca. 2 bis 3 Funktionen genutzt. Die am häufigsten genutzte Funktion ist das Erfragen einfacher Informationen, die aus dem Internet abgerufen werden, wie beispielsweise das Wetter, Sportergebnisse oder die nächste Tankstelle. Alexa, der Sprachassistent von Amazon, wird im Durchschnitt deutlich häufiger genutzt als die anderen Sprachassistenten. Dabei verwendet der durchschnittliche Alexa-Nutzer rund eine Funktion mehr als die Nutzer der anderen Systeme. Besonders beliebt ist das Abspielen von Musik, sei es über den Amazon-eigenen Dienst oder einen Drittanbieter.

Multihoming, also die parallele Nutzung mehrerer solcher Assistenten, ist die klare Ausnahme: 78% der Nutzer beschränken sich auf nur einen Sprachassistenten. Unter den 22% derjenigen, die mehrere Dienste parallel verwenden, finden sich sowohl Nutzer, die unterschiedliche Dienste für unterschiedliche Aufgaben wählen (etwa 7% der befragten Nutzer) sowie solche, die tatsächlich zumindest eine gleiche Funktion über mehrere Dienste ausführen (etwa 15% der befragten Nutzer). Die in anderen Endgeräten wie Autos, Kühlschränken oder Waschmaschinen integrierten Sprachassistenten, die oft von Drittanbietern bereitgestellt werden, spielen kaum eine Rolle in Deutschland.

Die festgestellten Nutzungsmuster, insbesondere das fehlende Multihoming sowie die enge Einbindung der populären Sprachassistenten in größere digitale Ökosysteme, unterstreichen die Relevanz der kontinuierlichen Beobachtung von digitalen Technologien. Das gilt insbesondere, wenn sie eine direkte Kundenschnittstelle und damit eine mögliche neue Gatekeeper-Position darstellen, wie es bei Sprachassistenten der Fall ist. Die große Verbreitung der vorinstallierten Systeme lässt erahnen, dass die breite Adoption schneller gehen könnte als bei bisherigen digitalen Technologien, sobald die Systeme die notwendige Attraktivität für Konsumenten haben. Deshalb ist eine vorausschauende Aufstellung der relevanten Behörden wichtig.



## Summary

Be it Amazon's Alexa, Apple's Siri, Google's Assistant, Microsoft's Cortana or Samsung's Bixby: Voice assistants are becoming more and more popular. According to the survey of 3,184 consumers in Germany, which was conducted for this discussion paper, 26% already used at least one such voice assistant at the end of 2018. The adoption rate is thus similar to that of smartphones about five years after the introduction of the first wave of devices.

However, the two technologies differ critically as regards the investment and effort a consumer has to take in order to use them. With around 85% of German consumers already owning or using at least one device with a pre-installed voice assistant, they do not have to spend additional money or even to install an additional app to start using a voice assistant. The majority of current smartphones and laptops feature pre-installed voice assistants. So called smart speakers have found their way into the homes of around 11% of the respondents.

Given the penetration of devices with pre-installed voice assistants, a rapid increase in the adoption rate is possible. In light of our survey, however, such an increase is unlikely in the near future. The vast majority of non-users have no intention of using one of the five most popular voice assistants within the next year. Current usage patterns also do not point to a great influence of the assistants. Consumers use voice assistants rarely and only for about 2 to 3 functions. The most frequently used one is to request simple information from the Internet, such as the weather, sports results or the nearest petrol station. Alexa, Amazon's voice assistant, is used much more frequently on average than the other systems. The average Alexa user also relies on approximately one function more than the users of Bixby, Cortana, Google Assistant and Siri. Music streaming is particularly popular with Alexa users, be it via Amazon's own service or a third-party provider.

Multi-homing, i.e. the parallel use of several such assistants, is the exception: 78% of users rely on one voice assistant. Among the 22% of those who multi-home, there are users who choose different services for different tasks (about 7% of the surveyed users) as well as those who actually perform at least one identical function over several services (about 15% of the surveyed users). The voice assistants integrated in other devices such as cars, refrigerators or washing machines, which are often provided by third-party providers, hardly play a role in Germany.

The usage patterns observed, especially the lack of multi-homing and the close integration of popular voice assistants into larger digital ecosystems, underline the importance of the continuous monitoring of digital technologies. This is particularly true if they represent a direct customer interface and thus a possible new gatekeeper, as is the case with voice assistants. The huge number of pre-installed systems suggests that broad adoption of voice assistants could be substantially faster than with existing digital technologies as soon as the systems have the necessary appeal for consumers. Competent authorities should therefore develop forward-looking policies with regard to this upcoming categories of services.

## 1 Einleitung

In den vergangenen vier bis fünf Jahren wurden Sprachassistenten wie die bekannten Dienste Alexa (Amazon), Bixby (Samsung), Cortana (Microsoft), Google Assistant (Google) und Siri (Apple) zusehends populärer.<sup>1</sup> Ihre Popularität kann auf die deutlichen technologischen Fortschritte zurückgeführt werden, die die Spracherkennung und -ausgabe in den letzten Jahren insbesondere aufgrund der Weiterentwicklung des Natural Language Processing (NLP) gemacht hat. Somit können heutige Systeme erheblich natürlicher auf Anfragen ihrer Nutzer antworten und bieten im Vergleich zu frühen sprachbasierten Benutzerschnittstellen einen deutlich erweiterten Funktionsumfang.

Typischerweise sind die genannten Systeme in modernen Endgeräten wie Smartphones, Tablets oder Desktop- bzw. Laptop-Computern vorinstalliert. Darüber hinaus finden sich die Systeme als integraler Bestandteil in sogenannten Smart-Speakern, die oftmals mit dem Zweck, einen Sprachassistenten im eigenen Zuhause ständig verfügbar zu haben, beworben werden. Trotz der vermutlich großen Verbreitung dieser Systeme aufgrund ihrer Vorinstallation auf verschiedenen beliebten Endgeräten, die verhältnismäßig kurzen Produktlebenszyklen unterliegen bzw. regelmäßig Aktualisierungen ihrer Software erfahren, gibt es bisher relativ wenige fundierte Daten zur tatsächlichen Nutzung von Sprachassistenten in Deutschland. Noch weniger Informationen finden sich zu den detaillierten Nutzungsmustern der Systeme oder den Gründen, aus welchen Konsumenten Sprachassistenten nicht mehr nutzen.

Auf diese Punkte geht der vorliegende Diskussionsbeitrag ein. Im Diskussionsbeitrag wird zum einen die Art und Weise der Nutzung von Sprachassistenten in Deutschland auf Basis einer Befragung von 3.184 Konsumenten betrachtet. Andererseits wird auf die Einstellungen der Konsumenten zu Sprachassistenten eingegangen. Aus diesen Informationen sowie einer umfangreichen Literaturrecherche zur Evolution und heutigen Funktionalitäten von Sprachassistenten werden dann mögliche Herausforderungen für Wettbewerb und Regulierung abgeleitet. Dabei wird den Themen Datenschutz und Sicherheitsaspekte bei Sprachassistenten besonderes Augenmerk geschenkt.

Der erste Teil (Kapitel 2) des Diskussionsbeitrags gibt einen Überblick darüber, was Sprachassistenten sind, aus welchen Komponenten sie sich zusammensetzen und welche Fähigkeiten sie besitzen. Ebenso umreißt der erste Teil die Themen Datenschutz und Sicherheitsaspekte von Sprachassistenten. Der zweite Teil (Kapitel 3 und 4) des Diskussionsbeitrags widmet sich den Ergebnissen der durchgeführten Befragung und reflektiert diese vor dem Hintergrund der vorhandenen Literatur zu Nutzungsmustern von Sprachassistenten. Im dritten Teil (Kapitel 5) werden mögliche wettbewerbliche und regulatorische Herausforderungen beschrieben. Der Diskussionsbeitrag schließt mit einem Fazit.

---

<sup>1</sup> Vgl. bspw. PwC (2017), PwC (2018).

## 2 Überblick zu Sprachschnittstellen und -assistenten

Bevor sich dieser Diskussionsbeitrag den Nutzungsmustern von Sprachassistenten in Deutschland widmet, gibt dieses Kapitel zunächst einen Überblick zu den grundsätzlichen Eigenschaften und Funktionen von Sprachassistenten. Anschließend werden wesentliche Aspekte in Bezug auf Verbraucherschutz bei diesen Systemen herausgegriffen und beschrieben. Das sind an dieser Stelle insbesondere Datenschutz und Sicherheitsaspekte der Systeme. Neben diesen Aspekten geht der Diskussionsbeitrag in Kapitel 5 auf die Möglichkeit eines neuen Gatekeepers sowie die bei Sprachassistenten besonders relevanten Netzwerk- und Feedbackeffekte ein. Diese haben ebenfalls Relevanz für den Verbraucherschutz, da sie ggf. den Auswahlraum der Konsumenten verkleinern. Ein Überblick zu bisher eingeführten und geplanten Sprachassistenten findet sich auch am Anfang in Abschnitt 3.1.

### 2.1 Definition und Abgrenzung verschiedener Typen von Sprachschnittstellen und -assistenten

Typischerweise funktioniert die Eingabe von Informationen und die Bedienung von informationstechnischen Systemen über mechanische Hilfsmittel wie eine Tastatur oder eine Computermouse in Kombination mit einer grafischen Benutzeroberfläche (Englisch: Graphical User Interface (GUI)). Alternativ kommen haptische Oberflächen wie z. B. ein berührungsempfindlicher Bildschirm (Touchscreen) zum Einsatz, die die mechanische Eingabe im GUI integrieren. Sprachschnittstellen, im Englischen Voice User Interface (VUI)<sup>2</sup> genannt, ermöglichen den Wegfall dieser mechanischen Hilfsmittel zur Bedienung des informationstechnischen Systems bzw. zur Eingabe und zum Abruf von Informationen.<sup>3</sup>

Während die Sprachschnittstelle einen wesentlichen Baustein von Sprachdialogsystemen (Spoken Dialogue Systems, SDSs) darstellt, gehören zu einem funktionierenden System immer noch weitere Komponenten (McTear 2002). Diese Komponenten ermöglichen typischerweise die folgenden Funktionen (Wyard et al. 1996):

- Speech recognition
- Language understanding
- Dialogue Management
- Communication with external systems

---

<sup>2</sup> Der Begriff wird in der Literatur nicht ganz einheitlich verwendet. Teilweise werden auch Gesamtsysteme, die in ihrer Funktionsweise Sprachassistenten ähneln, darunter zusammengefasst. Im Wesentlichen werden VUIs jedoch als das Interface und damit nur als ein Baustein von umfassenderen Systemen wie Sprachassistenten gesehen.

<sup>3</sup> Laut Wikipedia wird der Begriff „Stimmliche Benutzerschnittstelle (SBS)“ ebenfalls verwendet. Im Rahmen der Recherche für diesen Beitrag konnten hierfür jedoch keine Anhaltspunkte gefunden werden.

- Response generation
- Speech output

Welche einzelnen Komponenten dabei für welche Funktion eingesetzt werden, hängt von der spezifischen Umsetzung des jeweiligen Systems ab. McTear (2002) gibt dazu einen umfangreichen Überblick. Jokinen & McTear (2009), McTear, Callejas, & Griol (2016) und Olaso Fernández (2017) können als weitere aktuellere Quellen herangezogen werden, um einen Überblick zur Funktionsweise und zum Aufbau von Sprachdialogsystemen zu bekommen.

Tabelle 2-1 gibt einen Überblick zu verschiedenen Definitionen von Sprachschnittstellen und Sprachdialogsystemen. Die einfachsten Sprachdialogsysteme sind so genannte Interactive Voice Response (IVR) Systeme (Xu et al. 2010). Dieser Begriff bezeichnet eine Gruppe von Systemen, die automatisierte Abfragen wiedergeben und entweder gesprochene (einfache) Antworten oder zumeist im Tonwahlverfahren ausgewählte Antworten im Rahmen von Call-Center-Anwendungen verarbeiten. Die nächsthöhere Stufe von Systemen, die eine Spracheingabe und -ausgabe ermöglichen, werden als Speech Interfaces (Weinschenk & Barker 2000) bezeichnet. Sie verlagern im Wesentlichen die Spracheingabe von Befehlen, anstatt diese manuell einzugeben. Diese Systeme sind zumeist jedoch auch in der Lage, korrektive Nachfragen zu stellen, um nicht verstandene, unzutreffende oder anderweitig nicht bearbeitbare Befehle erneut zu erfragen oder Eingaben zu bestätigen. Sie bilden das Kernstück für aufwändigere und stärker an die natürliche Sprache angepasste sogenannte Conversational Interfaces (Zue & Glass 2000, McTear, Callejas, & Griol 2016). Diese Technologie stellt die Grundlage für Sprachassistenten wie die fünf hier betrachteten Systeme – Amazons Alexa, Apples Siri, Googles Assistant, Microsofts Cortana und Samsungs Bixby – dar. Während sich der Begriff des Conversational Interface schon länger etabliert hat, bestreiten Porcheron et al. (2018) auf Basis von 883 Aufnahmen von Anfragen an Alexa, dass diese Systeme in der Praxis tatsächlich als „conversational“ verwendet werden. Vielmehr handelt es sich nach dieser Analyse immer noch um eine sehr klare Input-Output-Logik, die Nutzer dieser Systeme verwenden. Die Ergebnisse anderer Forscher weisen in eine ähnliche Richtung, indem sie zeigen, dass Nutzer von Sprachassistenten zunächst lernen (müssen), wie sie erfolgsversprechende Anfragen formulieren, die von den jeweiligen Systemen korrekt erkannt und beantwortet werden (Myers et al. 2018, Luger & Sellen 2016, Rong et al. 2017, Chen & Wang 2018, Yarosh et al. 2018, Jiang, Jeng, & He 2013, Pelikan & Broth 2016, Beneteau et al. 2019, Reeves, Porcheron, & Fischer 2018, Sciuto et al. 2018).

Tabelle 2-1: Überblick zur Definition von Sprachschnittstellen bzw. Sprachdialogsystemen verschiedener Ausprägung in der Literatur

Quelle	Definition
Cohen, Giangola, & Balogh (2004)	„A voice user interface (VUI) is what a person interacts with when communicating with a spoken language application. The elements of a VUI include prompts, grammars, and dialog logic (also referred to as call flow). The prompts, or system messages, are all the recordings or synthesized speech played to the user during the dialog. Grammars define the possible things callers can say in response to each prompt. The system can understand only those words, sentences, or phrases that are included in the grammar. The dialog logic defines the actions taken by the system.” (p. 5)
Fraser (1998)	Interactive dialogue systems are “defined most generally as computer systems with which humans interact on a turn-by-turn basis.”
McTear, Callejas, & Griol (2016)	Conversational interfaces “enable people to interact with smart devices using spoken language in a natural way-just like engaging in a conversation with a person.” (p. 1)
Sciuto et al. (2018)	“A conversational user interface (CUI) is a system that relies on chat, supported by artificial intelligence, to have verbal interactions with end users.” (p. 1)
Zue & Glass (2000)	„[...] a new breed of human computer interfaces that combines several human language technologies to enable humans to converse with computers using spoken dialogue for information access, creation, and processing.” (p. 1166)
Baber (1993)	Digital Assist “,that uses input such as the user’s voice [...] and contextual information to provide assistance by answering questions in natural language, making recommendations and performing actions. (p. 223).
Azvine et al. (2000)	“The Intelligent Assistant (IA)integrated system of intelligent software agents that helps the user with communication, information and time management.”(p. 1)
Budzinski, Noskova, & Zhang (2018)	A DPA is an intelligent automated system that interacts with the user through a dialogue in natural language, and meanwhile applying third-party services to obtain information and perform various actions. (p. 4)
Hauswald et al. (2015)	An IPA is an application that uses input such as the user’s voice, vision (images) and contextual information to provide assistance by answering questions in natural language, making recommendations, and performing actions.” (p. 223).

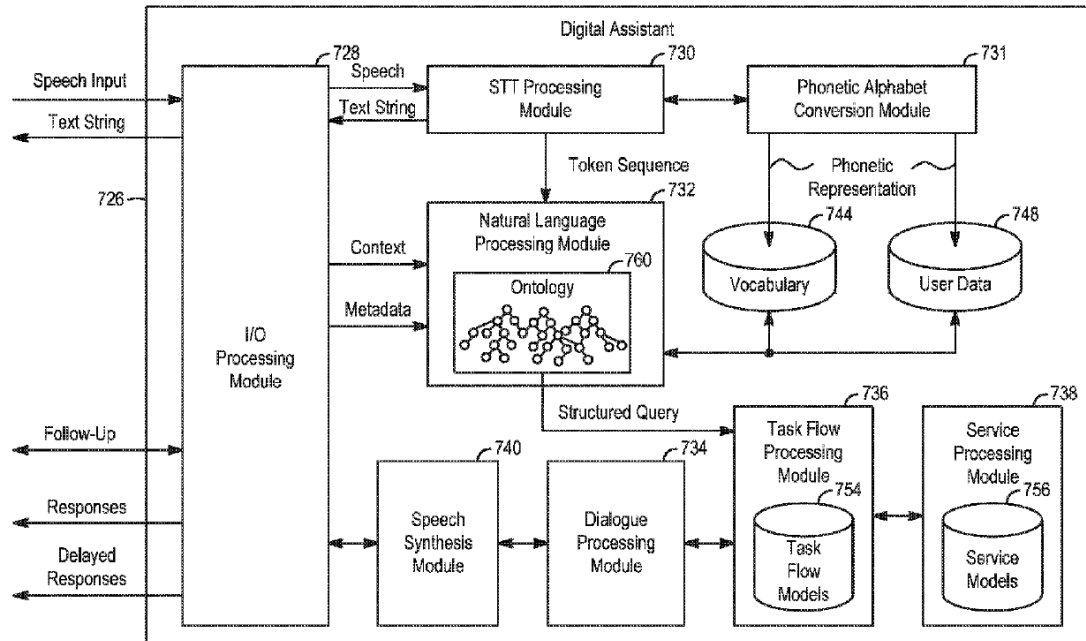
Quelle: Eigene Recherchen.

Die Definition von McTear, Callejas, & Griol (2016) weist auf den Unterschied zwischen der technischen Schnittstelle an sich und dem gesamten System eines intelligenten (smarten) Endgeräts oder Dienstes hin. So stellt in Bezug auf die hier betrachteten Sprachassistenten die technische Sprachschnittstelle ebenfalls einen wesentlichen Bestandteil des Dienstes dar.

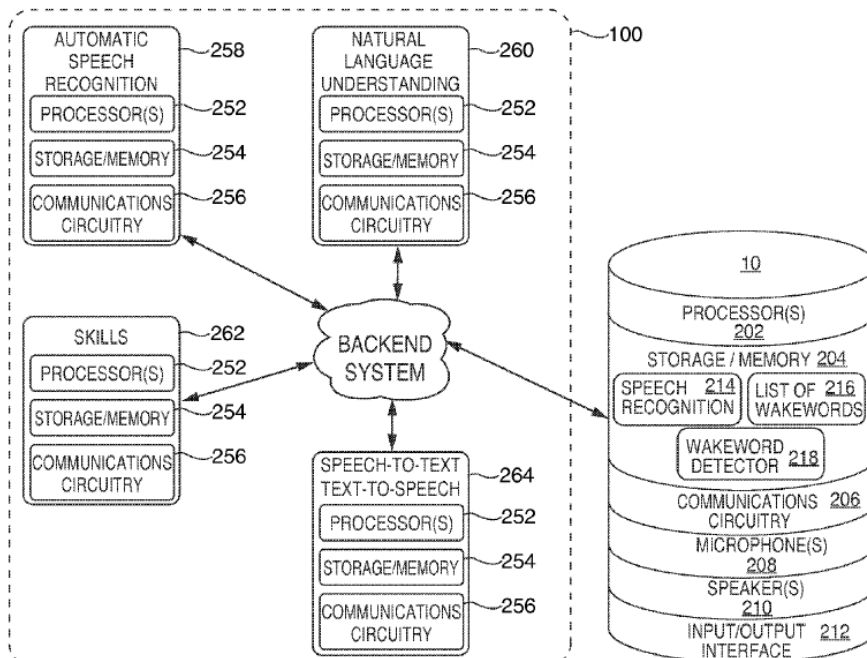
Der gesamte Dienst bzw. das vollständige Endgerät, so wie der Nutzer es verwendet, besteht jedoch aus zahlreichen weiteren Bausteinen, wie Abbildung 2-1 illustriert. Die Sprachschnittstelle bezeichnet ausschließlich das jeweilige Input/Output-Modul bzw. Interface.

Abbildung 2-1: Beispiele für verschiedene populäre Sprachassistenten aus Patentanmeldungen der Hersteller<sup>4</sup>

**Apple**

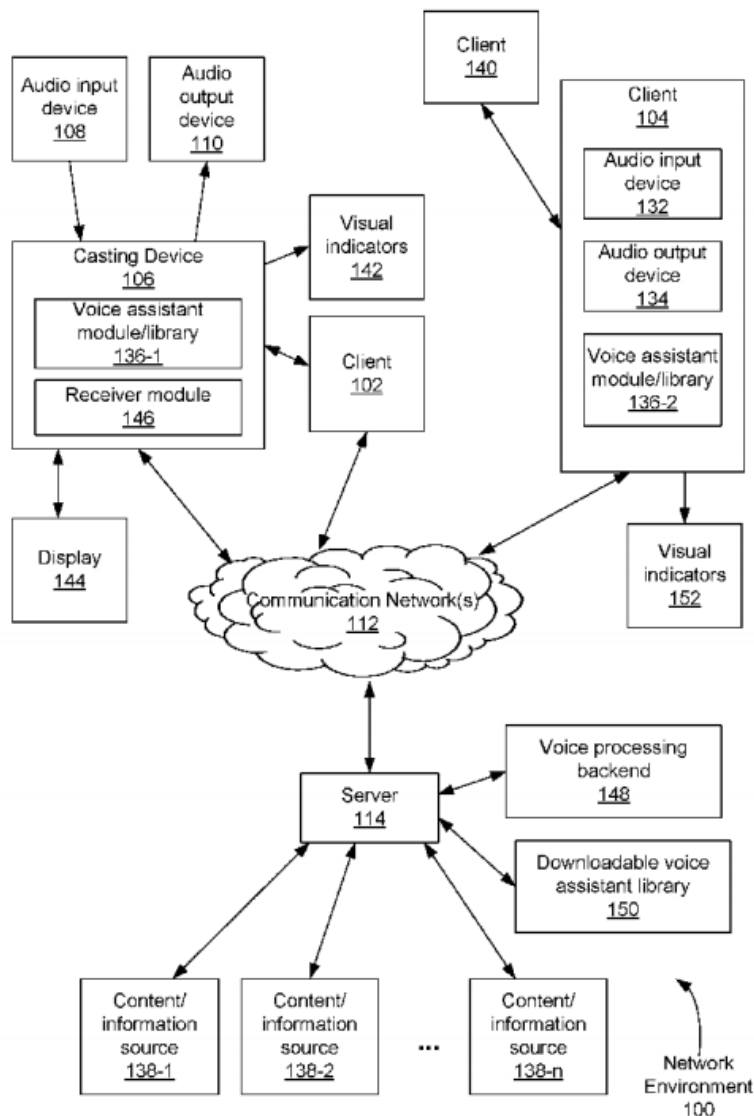


**Amazon**



<sup>4</sup> Für eine vollständige Beschreibung der Funktionsweise der einzelnen Komponenten wird auf die als Quellen angegebenen Patente verwiesen.

Google



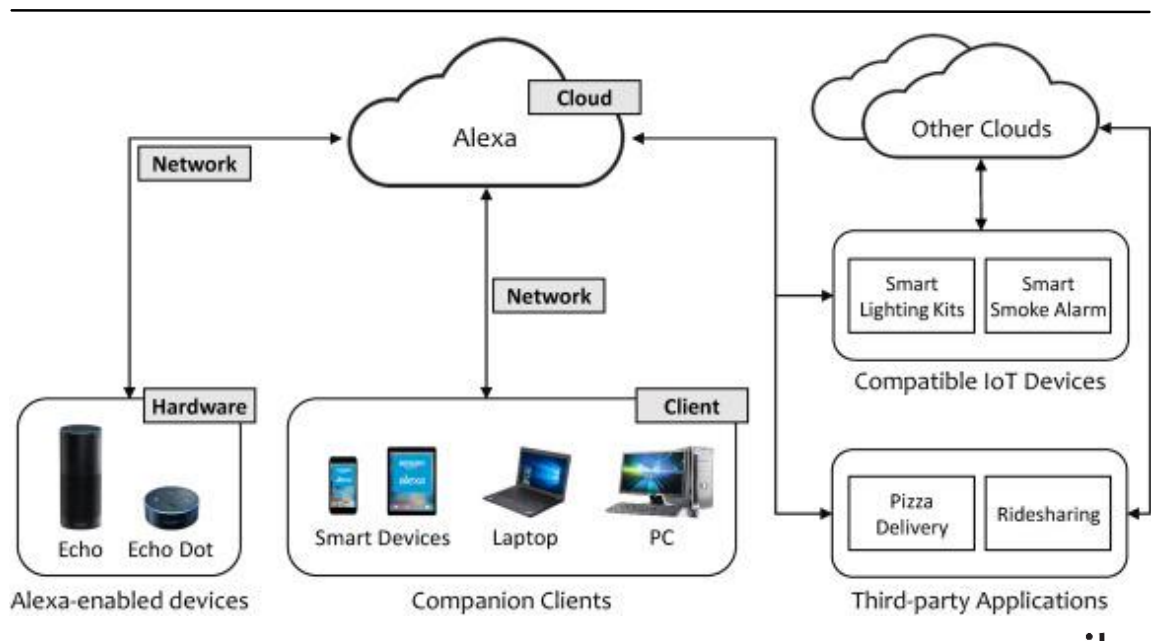
Quellen: US Patent Application Publication Pub. No. 2018/0330731 A1; US Patent No. 9,691,378 B1; International Publication Number WO 2017/197010 A1.

Insbesondere die Spracherkennung hat in den letzten Jahren von einem deutlichen Entwicklungsschub profitiert. So berichten schon Huang, Baker, & Reddy (2014), dass die in alten Systemen typischen Hidden Markov Models (HMMs) inzwischen durch Deep Neural Networks (DNNs) ersetzt wurden. Recurrent Neural Networks (RNNs) haben die sogenannten n-gram-Modelle abgelöst, die zur Ermittlung der Wahrscheinlichkeiten von Folgewörtern verwendet wurden. Ebenso wurden Conditional Random Fields (CRFs) durch neuronale Netze ersetzt, um Nutzerintentionen zu identifizieren (Sarikaya 2017). So wurden deutlich geringere Fehlerraten beim Erkennen von Sprach-eingaben möglich. Es ist zu erwarten, dass weitere Fortschritte bei neuronalen Netzen

und anderen Methoden zu spürbaren Verbesserungen bei der Sprach- und Intentionserkennung führen werden.

Wie oben gezeigt wurde, handelt es sich bei der Sprachschnittstelle nur um einen Baustein von Sprachassistenten. Doch auch diese sind zumeist in ein weiteres Ökosystem von kompatiblen Endgeräten sowie die sonstige Infrastruktur des Herstellers des Sprachassistenten und ausgewählter Drittanbieter eingebunden. Chung et al. (2017) illustrieren dies am Beispiel von Amazons Alexa.<sup>5</sup>

Abbildung 2-2: Das Amazon Alexa-Ökosystem



Quelle: Chung, Park, & Lee (2017) und Chung et al. (2017).

Heute existiert ein breites Spektrum an unterschiedlichen digitalen Sprachassistenten, die sich durch spezifische Charakteristika voneinander abgrenzen lassen. Knotte et al. (2019) haben eine Clusteranalyse durchgeführt, die vor allem auf der systematischen Auswertung von Literatur zu verschiedenen sowohl für die Forschung als auch für den kommerziellen Gebrauch entwickelten digitalen Assistenten beruht. Insgesamt wurden für den Zeitraum ab 2000 115 verschiedene digitale Assistenten identifiziert und untersucht. Insgesamt fanden die Autoren etwa 31 unterschiedliche Charakteristika, die die einzelnen Assistenten aufwiesen, die in 10 Dimensionen zusammengefasst wurden. Zusammenfassend konnten fünf Cluster identifiziert werden. Die nachfolgende Tabelle fasst die Ergebnisse der Untersuchung grob zusammen.

<sup>5</sup> Kapitel 5 geht auf die daraus resultierenden Netzwerk- und Feedbackeffekte ein.



Tabelle 2-2: Ergebnisse der Clusteranalyse von Knotte et al. (2019)

Characteristics	SPAs	Cluster				
		C1 (26)	C2 (19)	C3 (38)	C4 (15)	C5 (17)
<b>communication mode</b>						
text	18		x			
voice	23					x
vision	3					
text and vision	6					
voice and vision	57	x		x		
passive / observational	8				x	
<b>direction of explicit interaction</b>						
user-to-system	4					
system-to-user	18				x	
bidirectional	93	x	x	x		x
<b>query input</b>						
formal prompts	12					
natural language	83	x	x	x		x
sensor data	20				x	
<b>response output</b>						
vision	35		x		x	
voice	20					x
voice and vision	60	x		x		
<b>action</b>						
no service execution	65		x	x	x	
service execution	50	x				x
<b>assistance domain</b>						
general	45	x				x
specific	70		x	x	x	
<b>accepted commands</b>						
none	50		x	x	x	
manual data entry	17		x			
primitive commands	36	x				
compound commands	12					x
<b>adaptivity</b>						
static behavior	64		x	x	x	x
adaptive behavior	51	x				
<b>collective intelligence</b>						
no crowd engagement	93		x	x	x	x
crowd engagement	22	x				
<b>embodiment</b>						
none	30				x	
virtual character	14		x			
artificial voice	28	x				x
virtual character with voice	43			x		

Quelle: Eigene Darstellung basierend auf Knotte et al. (2019). SPA = Smart Personal Assistant; die Kreuze geben das jeweils dominierende Charakteristikum an.

Die sich ergebenden fünf Cluster sind: Adaptive Voice (Vision) Assistants (C1), Chatbot Assistants (C2), Embodied Virtual Assistants (C3), Passive Pervasive Assistants (C4) und Natural Conversation Assistants (C5).<sup>6</sup> Jeder von diesen weist ein anderes Set an

<sup>6</sup> Knotte et al. (2019) beschreibt die einzelnen Cluster wie folgt: Adaptive Voice (Vision) Assistants (C1): Dieses Cluster „contains SPAs that assist users mainly via speech and, optionally, also via optical sensors and visual output on a screen. [...] Therefore, these systems are capable of both understanding and responding in natural language and execute [...] services upon user requests. The vast majority [...] obtains knowledge models for general purposes, such as controlling smart household gadgets, retrieving mails or adding calendar entries. These knowledge models [...] are adaptive as they evolve over longer usage periods and, thus, provide higher service quality when used regularly.“ Chatbot Assistants (C2): Dieses Cluster „contains SPAs which mainly rely on text chat interaction to provide assistance services. This especially comprises chatbots, text-based conversational agents that are able to react to user input

unterschiedlichen dominierenden Charakteristiken auf, wie Tabelle 2-2 verdeutlicht. So nutzen digitale Assistenten, die zum Cluster 1 gehören, überwiegend eine audiovisuelle Kommunikationsform, sind bidirektional, reagieren auf natürliche Sprache, jedoch nur auf primitive Befehle usw. Knotte et al. (2019) sortieren hier vor allem die *General-Purpose-Assistenten* wie die bekannten Dienste Alexa, Bixby, Cortana, Google Assistant und Siri.<sup>7</sup> Dies sind vor allem allgemeine und auf eine breite Masse an unterschiedlichen Nutzergruppen ausgerichtete Dienste. Sie sind sehr Konsumenten-zentriert erstellt. Die Assistenten in den anderen drei Clustern wurden vor allem für einen bestimmten Zweck entwickelt. Sie sind zwar dem Endverbraucher typischerweise nicht direkt namentlich bekannt, aber in vielen Endgeräten als integrierte Assistenten verbaut. Hier spielen vor allem Assistenten in Kraftfahrzeugen und anderen Endgeräten eine wichtige Rolle. Einige bekannte Zulieferer sind hier Nuance Communications, SoundHound, IBM und Speechmatics. Diese entwickeln individuelle Sprachassistenten für Privat- und Geschäftskunden.

Im Mittelpunkt dieser Arbeit stehen die sogenannten General-Purpose-Assistenten. Doch auch diese lassen sich nochmal unterteilen. Ein weiterer Unterschied liegt an der jeweiligen Steuerung der Dienste. Viele werden lediglich über die Stimme gesteuert, einige können jedoch auch über berührungsempfindliche Bildschirme, Gesten und Emotionen gesteuert werden. Ähnliches gilt bei der Ergebnisausgabe. Hier kann es sich entweder um eine rein auditive Ausgabe handeln oder die Ausgabe wird vollständig oder in Teilen über grafische oder anderen Benutzeroberflächen erbracht bzw. unterstützt.

Ein weiterer offensichtlicher Punkt, der Sprachassistenten aus Sicht der Konsumenten voneinander unterscheidet, ist die Endgerätgebundenheit. So ist Siri eng an Endgeräte des Herstellers Apple gebunden, während bspw. Alexa von Amazon sowohl mit den verschiedenen Echo-Endgeräten als auch als Applikation für (fast) jedes beliebige Endgerät nutzbar ist.<sup>8</sup> Weiterhin gibt es einige Sprachassistenten, die als „eigene Marke“

---

based on semantic text analysis. Such systems are increasingly employed in first-level support service encounters as they are able to answer frequently asked questions and guide users through support processes.” Embodied Virtual Assistants (C3): Dieses Assistenten in diesem Cluster sind „human-like, virtual assistants. This is accomplished by both speech and visual output. Systems are mainly screen-based to present a virtual character (or avatar) with natural language speech, mimics and gestures to provide familiar interaction. Often, these assistants are designed for a special purpose such as e-learning [...]” Passive Pervasive Assistants (C4): Assistenten, die diesem Cluster zugeordnet werden können, „are designed to be as unobtrusive as possible. This means that users have few interactions with the system itself while it collects data from [...] sensors and infers and recommends suitable action.” Natural Conversation Assistants (C5): Assistenten, die diesem Cluster zugeordnet werden können, „can [...] be considered assistant for the ‘next generation of service encounters’. Focusing on speech interaction, [natural conversation assistants] aim to increase the similarity to human-to-human natural language interaction. They thus encompass more sophisticated speech recognition and spoken language understanding capabilities than any other class. [...] The primary design goal is to imitate human natural language interaction to provide a most natural and familiar interaction experience.”

7 Auch Bohm et al. (2017) verwenden die gleiche Bezeichnung für die Dienste der Endgeräte Amazon Echo und Google Home.

8 Microsoft bereitet anscheinend einen ähnlichen Schritt für sein System Cortana vor. Siehe [https://www.golem.de/news/cortana-beta-microsoft-bereitet-sprachassistenten-als-separate-app-vor-1907-142240.html?utm\\_source=nl.2019-07-01.html&utm\\_medium=e-mail&utm\\_campaign=golem.de-newsletter](https://www.golem.de/news/cortana-beta-microsoft-bereitet-sprachassistenten-als-separate-app-vor-1907-142240.html?utm_source=nl.2019-07-01.html&utm_medium=e-mail&utm_campaign=golem.de-newsletter) [Letzter Zugriff:01.07.2019].

positioniert werden. Dazu gehören die fünf hier betrachteten Systeme genauso wie der von der Deutschen Telekom geplante Magenta Assistent. Die oben erwähnten Systeme, die in verschiedenste Endgeräte integriert werden, sind dem Konsumenten zumeist nicht namentlich bekannt und werden nicht separat beworben.

Dies sind nur die augenscheinlichsten Charakterisierungen von Sprachassistenten. Weitere Klassifizierungen werden mit wachsender Verbreitung und erweitertem Funktionsumfang der Systeme vermutlich in Zukunft entstehen.

## 2.2 Momentaufnahme der Funktionen und Eingriffsmöglichkeiten für den Nutzer

Im Zentrum dieser Arbeit stehen vor allem die die General-Purpose-Assistenten, da eine komplette Marktabdeckung für den Zweck dieses Diskussionsbeitrags zu weit gehen würde. Namentlich werden hier die Assistenten Alexa des Herstellers Amazon, Bixby von Samsung, Cortana von Microsoft, Google Assistant von Google und Siri von Apple betrachtet, die in Deutschland große Bekanntheit genießen (PwC 2017).

Wie Knotte et al. (2019) bereits in ihrer Analyse zeigten, sind es diese Dienste, die bisher nur sehr rudimentäre Befehle befolgen können. Dennoch besitzen sie eine Reihe an Funktionen, die mit diesen Befehlen ausgeführt werden können. Eine eigene erste Recherche hat ergeben, dass etwa 33 unterschiedliche Funktionen genutzt werden können. Dabei besitzen nicht alle der betrachteten Dienste alle identifizierten Funktionen, wie Table 2-1 verdeutlicht. Die meisten Funktionen aus der unten aufgeführten Liste bietet der Google Assistant.

Table 2-1: Funktionen von Alexa, Bixby, Cortana, Google Assistant und Siri

	Alexa	Bixby	Cortana	Google Assistant	Siri
Anrufen	Ja	Ja	Ja	Ja	Ja
Auto	Ja	Nein	Ja	Ja	Ja
Automated Calls	Nein	Nein	Nein	Ja	Nein
Befehlketten (z.B. Quick commands)	Ja Routines	Ja Quick Commands	Nein	Ja Routines	Ja Shortcuts
Bilingual	Nein	Nein	Nein	Ja	Nein
Gerätekompatibilität	Ja (Smartphones, externe Geräte)	Derzeit gering, aber großes Potenzial	Ja	Ja	Nur Apple Produkte, manche Autos
Datenkontrolle	Ja	Können anschließend gelöscht werden	Ja	Ja	Nein, aber relativ hohe proklamierte Datensicherheit
Deutsche Sprache	Ja	Nein	Ja	Ja	Ja
Deutscher Markt	Ja	Ja	Ja	Ja	Ja
Empfehlungen	(Ja)	Ja	Nein	Ja	Ja
Erinnerungen	Ja	Ja	Ja	Ja	Ja

	Alexa	Bixby	Cortana	Google Assistant	Siri
Facial Recognition	Nur mit Zusatzgeräten oder Programmen	Nein	Nein	Nein	Nein
Folgefragen	Nein (begrenzt)	Ja	Nein (sehr begrenzt)	Ja	Nein
Fragen beantworten	Ja	Ja	Ja	Ja	Ja
Hands-free mode	Ja	Ja	Ja	Ja	Ja
Handy-kontrolle (z.B. App öffnen, Flugzeugmodus anschalten)	Begrenzt	Ja	Ja	Ja	Ja
Internetsuche	Ja	Ja	Ja	Ja	Ja
Kalender	Ja	Ja	Ja	Ja	Ja
Musik abspielen	Ja	Ja	Ja	Ja	Ja
Nachrichten/ News	Ja	Ja	Ja	Ja	Ja
Online Shoppen	Ja	Bilder werden an Amazon gelinkt, aber kein Kauf über Sprache möglich	Microsoft Edge browser hat Cortana integriert, sucht nach Coupons beim Shoppen	Indirekt (Anwendungen dritter)	Nein
Programme Drittanbieter	Skills	Ja	Cortana Skills	Ja	Shortcuts
Smart Home	Ja	Ja	Ja (sehr geringe Auswahl)	Ja	Ja
Sport Resultate	Ja	Ja	Ja	Ja	Ja
Sprechen	Ja	Ja	Ja	Ja	Ja
Taxi bestellen	Ja (Skill)	Ja	Uber, Ja	Ja	Ja
Text vorlesen und senden	SMS to Android	Ja	Ja (Email)	Ja	Ja
Umgebungswahrnehmung (Bild)	Nein (nur Echo Look)	Ja	Nein	Nein	Nein
Verkehr	Ja	Noch nicht	Ja	Ja	Ja
Video Kontrolle	Ja	Ja	Nein (vllt. mittlerweile möglich)	Ja (mit Anwendungen und Equipment)	Ja
Voice recognition	Ja (multi-user support)	announced	Ja	Ja (multi-user support)	Ja, kein multi-user support derzeit
Wecker	Ja	Ja	Ja	Ja	Ja
Voice Customisation	Ja (Language, Gender)	Ja (Gender, „Speaking Style“)	Ja (Language, Gender)	Ja (Language, Gender)	Ja, (Language, Gender, Accent)

Quelle: Eigene Darstellung der WIK-Recherche. Manche Funktionen nur in USA oder außerhalb Deutschlands verfügbar. Kein Vollständigkeitsanspruch. Stand: 18.03.2019.

Zusätzlich zu diesen integrierten Funktionen, die die einzelnen digitalen Assistenten zulassen, können einzelnen Assistenten auch zusätzliche Funktionen hinzugefügt werden. Diese oft als „Skills“ bzw. „Apps“ oder „Actions“ bezeichneten Zusatzfunktionen für die unterschiedlichsten Kategorien von Bildung über Gesundheit, Entertainment und Kommerz werden meist von Drittanbietern entwickelt und über die Assistenten angeboten. Vergleichbar sind diese mit den Apps und Apps-Stores bei Smartphones. Heute werden bereits mehrere Tausend derartiger Apps und Skills angeboten und die Zahl

wächst täglich. Laut Voicebot.ai stieg die Anzahl an Amazon Alexa Skills in Deutschland, den USA und dem Vereinigten Königreich im vergangenen Jahr rasant an. In den USA wurde ein Anstieg um 120% beobachtet. Aktuell existieren in den USA knapp 57.000 Skills. In Deutschland stieg die Anzahl um immerhin 150% auf etwa 8.000 und im Vereinigten Königreich wurde ein Anstieg von mehr als 200% beobachtet. Dies sind Zahlen von Januar 2019.<sup>9</sup> Weltweit sollen im Dezember 2018 etwa 80.000 Skills im Alexa-Katalog angeboten worden sein, während es im Januar 2016 gerade 130 Skills waren.<sup>10</sup> Ähnliches ermöglicht auch der Google Assistant durch die Google Actions. Dennoch ist der Angebotsumfang wesentlich geringer als der von Alexa. Selbst in den USA, dem bisher größten Markt, konnten im Januar nur etwas mehr als 4.000 Actions gezählt werden. Dies macht etwa 7,5% dessen aus, was Alexa an Drittanbieter-Applikationen anbietet.<sup>11</sup> Auf Deutschland bezogene Daten wurden noch nicht veröffentlicht; jedoch ist zu vermuten, dass die Werte bei Google Actions auch in Deutschland geringer ausfallen als bei Alexa Skills.

Die Assistenten bieten den Konsumenten eine Vielzahl an unterschiedlichen Eingriffsmöglichkeiten in die Einstellungen. Die folgende Tabelle 2-2 listet die Einstellungen auf, die durch den Nutzer der einzelnen Assistenten verändert werden können. Dabei hat der Konsument den Einfluss auf einzelne Designeinstellungen wie Stimme, Ton, Farbe etc., aber auch auf Einstellungen, die auf den Datenschutz und Austausch von Daten abzielen. Alle betrachteten Assistenten ermöglichen es den Nutzern festzulegen, welche Sprache und Stimme der Assistent verwenden soll. Ferner können die Nutzer bei allen fünf Assistenten bestimmen, mit welchen Geräten oder Diensten eine Verbindung eingegangen werden soll.

---

<sup>9</sup> <https://voicebot.ai/2019/01/02/amazon-alexa-skill-counts-rise-rapidly-in-the-u-s-u-k-germany-france-japan-canada-and-australia/> [Letzter Zugriff: 02.07.2019].

<sup>10</sup> <https://www.statista.com/statistics/912856/amazon-alexa-skills-growth/> [Letzter Zugriff: 02.07.2019].

<sup>11</sup> <https://voicebot.ai/2019/02/15/google-assistant-actions-total-4253-in-january-2019-up-2-5x-in-past-year-but-7-5-the-total-number-alexa-skills-in-u-s/> [Letzter Zugriff: 02.07.2017].

Table 2-2: Änderbare Einstellungen

Änderbare Einstellungen					
	Alexa	Google Assistant	Siri	Cortana	Bixby
Aktivitätsempfehlungen				x	
Aktivitätsverlauf				x	
Animation beim Öffnen					x
Ansprache/Aktivierungswort	x	x	x	x	
Audio-Steuer-elemente (Lautstärke, Bass etc.)	x				
Auto-Updates					x
Bitte nicht stören	x				
Bluetooth-Geräte	x				
Browserverlauf				x	
Cloudsuche				x	
Datenzugriff auf andere installierte Apps (an/aus)			x		
Dialogverlauf	x				
Echo-Fernbedienung	x				
Einstellung beim Sperrbildschirm			x	x	x
Farbe					x
Geräteverlauf	x			x	
Gesprächseinstellungen	x	x		x	x
Kommunikation (E-Mail, Kontakte, Kalender, Nachrichten, etc.)	x	x		x	x
Kurzbefehle			x		
Maßeinheiten	x				
Notifikationen (Marketing etc.)					x
Password					x
Registrierung/Profile	x		x		x
Telefon-Benachrichtigungen				x	
SafeSearch				x	
Sprache	x	x	x	x	x
Spracheinkauf	x				
Sprachfeedback (an/aus)			x		
Standort	x	x		x	
Stimme		x	x	x	x
Tastenkombination				x	
Ton	x				
Update Plugins					x
Verbindung mit Geräten und/oder Diensten	x	x	x	x	x
WLAN-Einstellungen	x				
Zeitzone	x				

Quelle: Eigene Darstellung der WIK Recherche. Kein Vollständigkeitsanspruch. Stand: April 2019.

Auffällig ist, dass nicht alle identifizierten datenschutzrelevanten Funktionen bei allen fünf Assistenten im gleichen Umfang vorhanden sind. Das ist interessant, da Datenschutz einer der wesentlichen Aspekte des Verbraucherschutzes in Bezug auf Sprachassistenten darstellt. Der nächste Abschnitt widmet sich diesen Fragestellungen im Detail.

## 2.3 Sprachassistenten und Verbraucherschutz

Wie in Abschnitt 2.1 bereits dargestellt kommen mit einer neuen Art von Benutzeroberfläche (Sprachdialogsysteme) und ihrer Integration in neue Dienste (Sprachassistenten), die über verschiedene Endgeräte zur Verfügung stehen, neue Herausforderungen in der Bedienung auf Konsumenten zu. Gleichzeitig bringen Sprachassistenten neue Herausforderungen für den Verbraucherschutz mit sich. Darunter spielt der Datenschutz eine besondere Rolle, da die hier betrachteten Systeme ständig auf ein sogenanntes Wake-Word<sup>12</sup> warten, welches die Aufnahme und Verarbeitung von Befehlen aktiviert.<sup>13</sup> Zahlreiche Berichte zeigen, dass diese Technologie nicht immer vollständig im Sinne der Nutzer funktioniert. Weiterhin entstehen neue Herausforderungen, die die allgemeine Sicherheit der neuen Benutzeroberfläche betreffen. Anders als die manuelle Bedienung bedingt die Bedienung über eine Sprachschnittstelle nicht zwingend eine sichtbare Eingabe von Befehlen, sondern kann auch durch von Menschen als undefinierbare Geräusche oder Krach wahrgenommene Töne bzw. Kombinationen von Tönen außerhalb des menschlichen Hörspektrums gesteuert werden.

### 2.3.1 Datenschutz

Wie oben schon kurz eingeführt, werden die hier betrachteten Sprachassistenten typischerweise durch ein Wake-Word aktiviert.<sup>14</sup> Das bedeutet jedoch nicht, dass ausschließlich die nach dem Wake-Word wahrgenommenen Audioinhalte gespeichert werden. So können Echo-Endgeräte<sup>15</sup> ebenfalls kontinuierlich jeweils bestimmte Intervalle der Audioinhalte im Raum, in welchem sie platziert wurden, aufnehmen und lokal puffern, bis das Wake-Word erkannt wird. Diese Audioinhalte können dann zusammen mit der Audioaufnahme der Audioinhalte, die nach dem Wake-Word bis zum (vom Endgerät wahrgenommenen) Ende der Frage oder des Befehls bzw. innerhalb einer festgelegten Zeit nach dem Wake-Word aufgenommen wurde, zur Interpretation an das Backend des Dienstes in der Cloud gesendet werden.<sup>16</sup> Solange kein Wake-Word erkannt wird, werden die lokal gepufferten Informationen gelöscht.<sup>17</sup> Der Google Assistant überträgt ebenfalls einige 100 Millisekunden des vor dem Wake-Word aufgenommenen Audioinhalts sowie eine ähnlich lange Aufnahme der Geräusche nach dem wahrgenommenen Ende der Frage bzw. des Befehls an das Backend<sup>18</sup> des Dienstes.<sup>19</sup> Apples Siri ver-

---

<sup>12</sup> Das Wake-Word kann alternativ auch als Hotword oder Trigger-Word bezeichnet werden. Bekannte Wake-Words sind „Alexa“ für Amazons Alexa-System, „Hey, Siri“ für Apples Siri-System oder, „OK, Google“ für Googles Assistant.

<sup>13</sup> Typischerweise erfolgt die Spracherkennung des Wake-Word vollständig lokal auf dem jeweiligen Endgerät (Bisher Tarakji et al. 2018).

<sup>14</sup> Es gibt auch alternative Wege, um die Sprachassistenten zu aktivieren. Diese alternativen Wege bedingen aber zumeist eine manuelle Interaktion mit dem Endgerät und stehen somit der Idee eines (reinen) Sprachdialogsystems entgegen.

<sup>15</sup> Amazon Echo, Amazon Echo Dot, Amazon Echo Show, etc.

<sup>16</sup> Vgl. hierzu US Patent, Patent No. US 10,109,294 B1 zu „Adaptive Echo Cancellation“ beantragt von Amazon Inc. – insbesondere Spalte 3 Zeile 61ff; ähnliche Informationen finden sich in US Patent, Patent No. US 9,691,371 B1 zu Methods and devices for selectively ignoring captured audio data beantragt von Amazon Inc.

<sup>17</sup> Vgl. *ibid.* Spalte 4 Zeile 6ff.

<sup>18</sup> Vgl. hierzu US Patent, Patent No. 8,775,191 B1 zu „Efficient utterance-specific endpointer triggering for always-on hotwording“ beantragt von Google Inc.

wendet Kontextinformationen in Form von bekannten Präferenzen des Nutzers, Informationen über den aktuellen Zustand der verwendeten Hard- und Software, Informationen der vorhandenen Sensoren, die vor, während oder eine kurze Zeit nach der durch das Wake-Word gestarteten Interaktion wahrgenommen werden, ebenso wie vorherige Interaktionen zwischen dem Nutzer und dem System sowie weitere Informationen.<sup>20</sup> Die eigentliche Übertragung von Audioeingaben erfolgt jedoch erst, sobald Siri das vom Nutzer voreingestellte Wake-Word wahrnimmt.<sup>21</sup>

Dieses Puffern von Informationen und im Falle der Aktivierung durch das Wake-Word Übertragen der Audioinhalte an das Backend des Dienstes dient der Verbesserung der Antwort bzw. Korrektheit der ausgewählten Aktion. Das Ziel ist es, eine möglichst kontextspezifische Antwort geben oder Aktion ausführen zu können.<sup>22</sup> Die Fähigkeit, den Kontext der Interaktion zu erkennen und daraus Informationen über die Intention des Nutzers zu generieren, ist eine kritische Eigenschaft von Sprachassistenten.<sup>23</sup> Sie leistet einen entscheidenden Beitrag zur wahrgenommenen Qualität des Dienstes. Gleichzeitig bedingt das Einbeziehen des Kontexts und ggf. der Nutzungshistorie in die probabilistische Abschätzung der Nutzerintentionen, dass zusätzliche Daten gesammelt und ausgewertet werden. Dabei stellen sowohl die Übertragungsgeschwindigkeit als auch die Rechenleistung der heutigen Systeme eine gewisse Begrenzung des Umfangs der Datenmenge dar, die für eine bestimmte Anfrage verarbeitet werden kann, ohne dass die Wartezeit auf eine Reaktion unverhältnismäßig lang wird und zu Unzufriedenheit bei den Nutzern führt (López-Cózar et al. 2003).

Laut Knotte et al. (2018) wird in der Wissenschaft oft die Meinung vertreten, dass „context-awareness“, d.h. die Identifikation, Verarbeitung und Nutzung von kontextuellen Informationen wichtige Funktionen sind, um Sprachassistenten und andere digitale Assistenten überhaupt als „smart“ oder „intelligent“ klassifizieren zu können. Je mehr Information den Systemen bereitgestellt wird, desto besser können sie ihre Assistenzfunktion ausüben und möglicherweise auf Veränderungen reagieren und das Verhalten ihrer Nutzern antizipieren. Dies bedeutet aber auch, dass die Menge an Daten, die erfasst und übertragen wird, zusätzlich steigt, was bei vielen Nutzern Bedenken hinsichtlich Datenschutz und Sicherheit auslöst.

- 
- 19** Die neueste Generation des Google Assistant, die auf den kommenden Pixel 3 Smartphones erhältlich sein wird, überträgt laut Google keine Daten mehr an das Backend in der Cloud, da die Prozesse vollständig lokal auf dem Smartphone ablaufen. Vgl. Google Blog vom 7. Mai 2019 unter <https://blog.google/products/assistant/next-generation-google-assistant-io/> [Letzter Zugriff: 02.07.2019].
- 20** Vgl. Australian Patent, Patent No. AU 2017210578 B2 zu „Voice trigger for a digital assistant“ beantragt von Apple Inc. insbesondere Absatz 0074.
- 21** Vgl. US Patent, Patent No. US 2014/0222436 A1 zu „Voice trigger for a digital assistant“ beantragt von Apple Inc.. In diesem Patentantrag wird ebenfalls erläutert, dass die integrierten Lärm-, und Trigger-Detektoren nicht kontinuierlich Audiosignale aufnehmen, sondern in einem engen Rhythmus an und ausgeschaltet werden. Dies mindert die Ergebnisqualität der Detektion kaum, reduziert den Energieverbrauch jedoch deutlich.
- 22** Tatsächlich verbessern Informationen über den Kontext die Qualität der Antworten und Aktionen von Sprachdialogsystemen. Vgl. hierzu Chen et al. (2017), McTear (2002) oder Leong et al. (2005).
- 23** Schon in frühen Sprachdialogsystemen spielt das Erkennen der Nutzerintention eine wesentliche Rolle. Vgl. hierzu McTear, Callejas, & Griol (2016) oder McTear (2002).



### 2.3.2 Sicherheitsaspekte von Sprachassistenten

Eine neue Technologie, für welche die Erfahrungswerte fehlen, ist immer angreifbarer als Technologien, mit denen sowohl die Entwickler als auch die Benutzer schon lange vertraut sind. Das gilt insbesondere, wenn die Benutzeroberfläche relativ leicht zugänglich ist bzw. die Notifikationsfunktionen der Geräte und Dienste eingeschränkt sind. Diese Punkte sind bei Sprachassistenten alle erfüllt. Es handelt sich um eine Technologie, die erst seit wenigen Jahren wirklich auf dem Massenmarkt angekommen ist. Obwohl die Entwicklung der Systeme, die wir heute kennen, schon vor etwa 40 Jahren begann, hat man sich bisher anscheinend stärker darauf konzentriert, wie man die komplexen Prozesse der Spracherkennung und die probabilistische Antizipation der Intentionen des Nutzers zum Funktionieren bringt, als sich mit der Angreifbarkeit der Benutzeroberfläche an sich auseinanderzusetzen. Bei den Nutzern überwiegt anscheinend ebenfalls eher die Euphorie, Sprachbefehle umfangreich nutzen zu können, als die Fokussierung auf mögliche Risiken, die mit der neuen Benutzeroberfläche einhergehen.

Während alle Systeme, die mit dem Internet verbunden sind, einer gewissen Gefahr von Cyberattacken ausgesetzt sind, unterscheiden sich Sprachassistenten, gerade wenn sie in Form eines Smart Speakers oft unbeobachtet sind, ganz grundsätzlich von anderen Diensten und Endgeräten. Im Gegensatz zu diesen ist bei Sprachassistenten kein physischer Kontakt zu einer Tastatur oder Maus notwendig, um sie zu bedienen. Tonsignale reichen aus. Diese müssen weder die Form von Sprache haben, die für den Menschen verständlich ist, noch müssen diese Tonsignale innerhalb des vom menschlichen Gehör wahrnehmbaren Frequenzbereichs liegen. Weiterhin fehlt den Systemen zumeist der Bildschirm, der im lange erlernten Interaktionsverhalten von Menschen und Computern eine zentrale Rolle bei der Kontrolle der tatsächlich ausgeführten Aktionen und Programme spielt. So ist die Wahrscheinlichkeit, dass dem Nutzer eine gefälschte Webseite anhand ihrer visuellen Merkmale auffällt, höher als die Wahrscheinlichkeit, dass dies anhand einer ausschließlich über Sprachausgabe präsentierten Passwortabfrage geschieht. Letztlich ist die Spracherkennung trotz enormer Fortschritte in den letzten Jahren immer noch weit von der Genauigkeit der Eingabe von Befehlen über die Tastatur entfernt.<sup>24</sup>

Bispham, Agrafiotis, & Goldsmith (2018) geben einen Überblick zu möglichen Attacken über die Sprachschnittstelle von Sprachassistenten und entwickeln eine Taxonomie möglicher Angriffe. Sie unterscheiden zwischen offenen Angriffen z. B. durch hörbare Sprache und verdeckte Angriffe, die durch nicht sinnvolle Laute, Sprache im Umfeld der Interaktion mit dem Assistenten, die vom System missinterpretiert wird, Stille, Geräusche oder Musik durchgeführt werden können. Alle diese Angriffe wurden schon auf verschiedene Weise in der Literatur gezeigt. Die folgende Tabelle gibt einen Überblick zu einer Auswahl der bisher erfolgreichen Angriffe auf Sprachassistenten.

---

<sup>24</sup> Das liegt einerseits an der Technologie, andererseits aber an der hohen natürlich vorkommenden Fehlerrate in der menschlichen Sprache. Vgl. hierzu bspw. Kumar et al. (2018); Zhang et al. (2019).

Tabelle 2-3: Übersicht zu dokumentierten erfolgreichen Angriffen auf Sprachassistenten über die Sprachschnittstelle

Quelle	Kurze Beschreibung des Angriffs und seiner Auswirkungen
Sophos (2017) <sup>25</sup>	Im Jahr 2017 veröffentlichte Burger King eine Werbung, in der einer der Schauspieler mit dem Satz „OK Google, what is the Whopper burger?“ den Google Assistant aktivierte, der darauf die volle Liste der im Whopper Burger enthaltenden Zutaten vorlas. Die Liste stammte aus einem Wikipedia-Eintrag. Nach der Veröffentlichung fing Individuen an, diesen Artikel und die Liste an Zutaten umzuschreiben bis sowohl Google als auch Wikipedia dies stoppten.
Forbes Magazine (2016) <sup>26</sup>	Im Jahr 2016 berichtete das Forbes Magazine von einer Situation, in der ein Mann die abgeschlossene Tür eines Hauses öffnen konnte, indem er rief „Hey Siri, unlock the front door“. Dies war möglich, weil der Hausbesitzer den Sprachassistenten so platzierte, dass dieser von außen gesehen werden konnte und den Ruf aufnahm. Hoy (2018) wies jedoch darauf hin, dass Apple an einem System zur Stimmenerkennung (und damit zur Verifikation der befugten Nutzer des jeweiligen Systems) arbeite, um solche Vorfälle zu verhindern.
Carlini et al. (2016) Bispham, Agrafiotis, & Goldsmith (2019); Zhang et al. (2017); Song & Mittal (2017)	Die Wissenschaftler testeten die Fähigkeit des Sprachassistenten, versteckte Spracheingaben zu identifizieren. Versteckte Spracheingaben sind solche, die für den Menschen selbst entweder schwer oder gar nicht zu verstehen sind, für Maschinen jedoch verständlich sein können, oder grundsätzlich eine andere Bedeutung für Menschen als für die Sprachassistenten haben. Die Studien belegten, dass für Menschen schlecht oder gar nicht verständliche Tonsequenzen, Sprachassistentensysteme dazu bringen können, verschiedenen Aktionen auszuführen.
Diao et al. (2014); Alepis & Patsakis (2017)	Die Autoren zeigten den Einfluss, den eine Malware haben kann, die keine Einwilligungen seitens des Endverbrauches benötigt, um Aktionen auf dem betroffenen Endgerät auszuführen (zero permission apps). Solche Apps können als normale Apps getarnt sei, die jedoch Module enthalten, die einen Angriff ausüben können. Da sich diese Apps als normale Applikationen tarnen, können sie durch den Einwilligungsmechanismus des jeweiligen Betriebssystems fallen und werden daher oft als ungefährlich eingestuft. In Situationen, in denen das Telefon oder der Smart-Speaker nicht von ihrem eigentlichen Nutzer verwendet werden, spielen diese Malware-Apps eine Audiodatei ab, die den oft integrierten Text-to-Speech-Dienst verwendet. Nach dem die mit Hilfe der am Endgerät befindlichen Sensoren die Malware analysiert, ob eine Aktion ihrerseits unentdeckt bleibt, führt die App den vorprogrammierten Angriff auf das System durch.
Kumar et al. (2018); Zhang et al. (2019)	Die Interpretationsungenauigkeit des Sprachprozessors von Sprachassistenten, die einerseits durch die Fehlerrate der Technologie, aber andererseits durch die Fehlerrate der Benutzer in der Aussprache entsteht, kann ausgenutzt werden, indem Skills mit Namen im App-Store des Sprachassistenten eingestellt werden, die in ihrem Namen denen anderer Anbieter stark genug ähneln, um „versehentlich“ aktiviert zu werden. Diese Apps können dann eine Malware enthalten, die Angriffe auf das betroffene Endgerät ausführen können, oder das Ziel haben, Kombinationen aus Benutzernamen und Passwörtern abzufischen. Wissenschaftler haben festgestellt, dass heute eine nicht unwesentliche Anzahl an Skills bzw. Apps in den App-Stores verschiedener Sprachassistentensysteme existieren, die anfällig für diese Art von Angriffen sind.

Quelle: Bispham, Agrafiotis, & Goldsmith (2018), sowie eigene Recherchen.

<sup>25</sup> <https://nakedsecurity.sophos.com/2017/04/18/burger-king-triggers-ok-google-devices-with-tv-ad/> [Letzter Zugriff: 27.06.2019].

<sup>26</sup> <https://www.forbes.com/sites/aarontilley/2016/09/21/apple-homekit-siri-security/> [Letzter Zugriff: 27.06.2019].

Die hier gesammelten Beispiele für Angriffe auf Sprachassistenten machen deutlich, dass es noch einiger Anstrengungen sowohl auf Seiten der Hersteller als auch auf Seiten der Nutzer bedarf, um die Sicherheit der Systeme zu gewährleisten. Die dringendsten Herausforderungen betreffen augenscheinlich die Sicherung der Systeme gegen Befehle, die nicht von autorisierten Nutzern stammen sowie die Verbesserung der Spracherkennung und Verhinderung von Aktionsausführungen auf Basis von nicht-sprachlichen Audiosignalen.

### 3 Nutzungsintensität und Nutzungsmuster von Sprachassistenten in Deutschland

Das nachfolgende Kapitel geht auf die aktuelle Nutzung von Sprachassistenten ein. Im Wesentlichen gibt dieses Kapitel einen Überblick zu den Fragen, inwieweit Sprachassistenten derzeit verbreitet sind, wie sie genutzt werden und wofür. Die Einblicke, die in diesem Kapitel präsentiert werden, basieren auf den Ergebnissen einer Online-Befragung des WIK aus dem Jahr 2018 zur Nutzung der Sprachassistenten Alexa, Bixby, Cortana, Google Assistant und Siri in Deutschland.<sup>27</sup> Um die Nutzung innerhalb Deutschlands ins Verhältnis zu stellen, werden ebenfalls internationale Studien diskutiert, die ähnliche Befragungen für ein Set an verschiedenen Ländern durchgeführt haben.

#### 3.1 Marktdurchdringung und Marktanteile

Obwohl, wie in Kapitel 2 angedeutet, Endgeräte, die über Spracheingabe gesteuert werden können, keine Neuerung des 21. Jahrhunderts sind, sondern bereits vorher existierten<sup>28</sup>, ist vor allem in den vergangenen vier Jahren ein deutlicher Anstieg in Angebot und Nachfrage der sogenannten General-Purpose-Assistenten zu beobachten, die auf die allgemeinen Bedürfnisse von Endkonsumenten ausgerichtet sind. Ein Grund hierfür ist vor allem die starke Verbesserung der technischen Möglichkeiten in der Stimmenerkennung und -bewertung, dem Kontextbewusstsein und Dialogmanagement (siehe Kapitel 2).<sup>29</sup> Schätzungen zufolge soll die Anzahl der Nutzer von digitalen Assistenten in den folgenden Jahren stark ansteigen. Erwartet wird ein Anstieg von 390 Millionen in 2015 auf 1,8 Milliarden Nutzer weltweit in 2021 (Knote et al. 2018, Knote et al. 2019).<sup>30</sup> Bis 2020 soll das Marktvolumen etwa 9 Milliarden Dollar betragen (Clark et al. 2019).

---

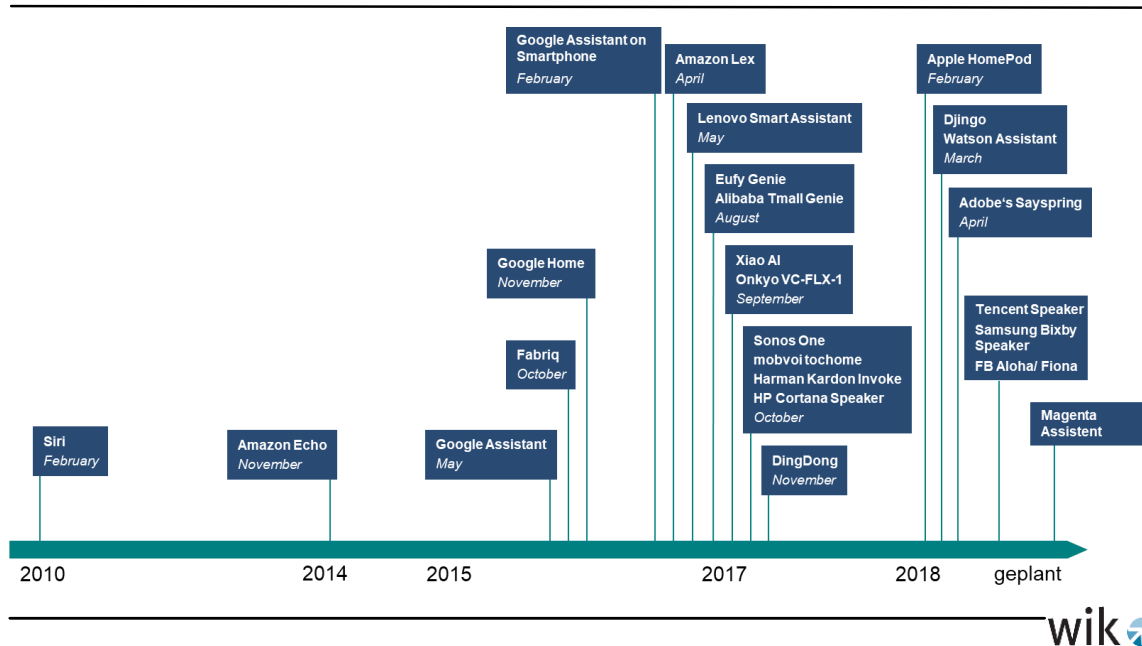
<sup>27</sup> Im Anhang findet sich eine ausführliche Beschreibung der angewandten Methodik.

<sup>28</sup> Siehe hierzu die Ausführungen in Arcep (2018).

<sup>29</sup> Dies fällt auch Konsumenten auf: 50% bis 60% der Konsumenten in den USA ist eine eindeutige Verbesserung der Systeme aufgefallen (Kinsella & Mutchler 2018).

<sup>30</sup> Zitiert nach <https://www.tractica.com/newsroom/press-releases/the-virtual-digital-assistant-market-will-reach-15-8-billion-worldwide-by-2021/> [Letzter Zugriff: 01.07.2019].

Abbildung 3-1: Zeitliche Entwicklung von intelligenten Sprachschnittstellen



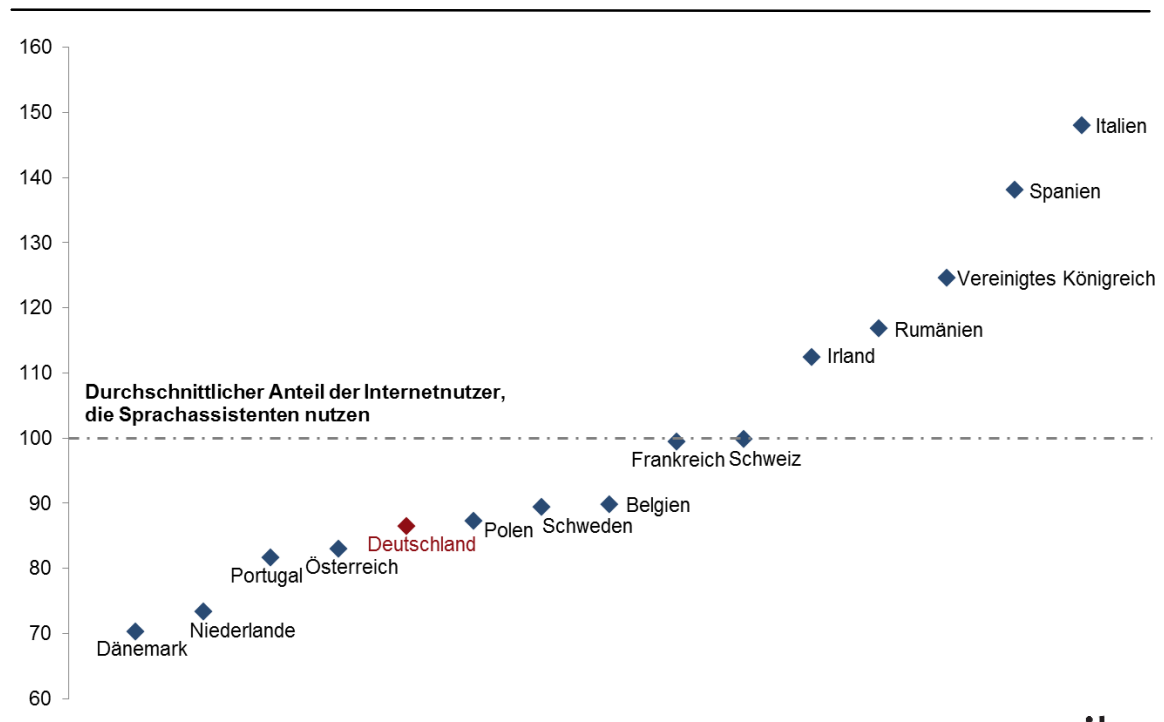
Quelle: Eigene Darstellung.

Die Online-Befragung des WIK zeigt, dass nur 30% der Befragten ein Endgerät mit einer Spracheingabe in 2018 bedient haben. Unter diesen Befragten nutzen die meisten die neue Benutzerschnittstelle, um ihr Smartphone zu bedienen.<sup>31</sup> Grundsätzlich liegt Deutschland jedoch im intra-europäischen Vergleich in der Nutzung von Sprachassistenten stark zurück. Im Vergleich der Internetnutzer in den Ländern in Abbildung 3-2, die Sprachassistenten im vierten Quartal 2018 genutzt haben, liegt Deutschland deutlich unter dem europäischen Durchschnitt. Besonders weit fortgeschritten in der Adoption sind vor allem Länder wie Italien, Spanien und das Vereinigte Königreich. Asiatische Länder wie China, Indien, Indonesien liegen weit über dem europäischen Durchschnitt. Diese Länder erreichen einen Index-Wert von über 200. Aber auch die USA liegen mit einem Index-Wert von etwa 150 weit vor Europa und Deutschland.<sup>32</sup>

<sup>31</sup> Die große Mehrheit der Endgeräte wurde dabei mit einer im Endgerät eingebauten Sprachsteuerung bedient (59%). 43% gaben hingegen an, das Endgerät durch die Verknüpfung mit einem externen Sprachassistenten bedient zu haben.

<sup>32</sup> <https://www.globalwebindex.com/> [Letzter Zugriff: 09.04.2019].

Abbildung 3-2: Anteil der Internetnutzer, die einen Sprachassistenten genutzt haben – Europa (Q4 2018)

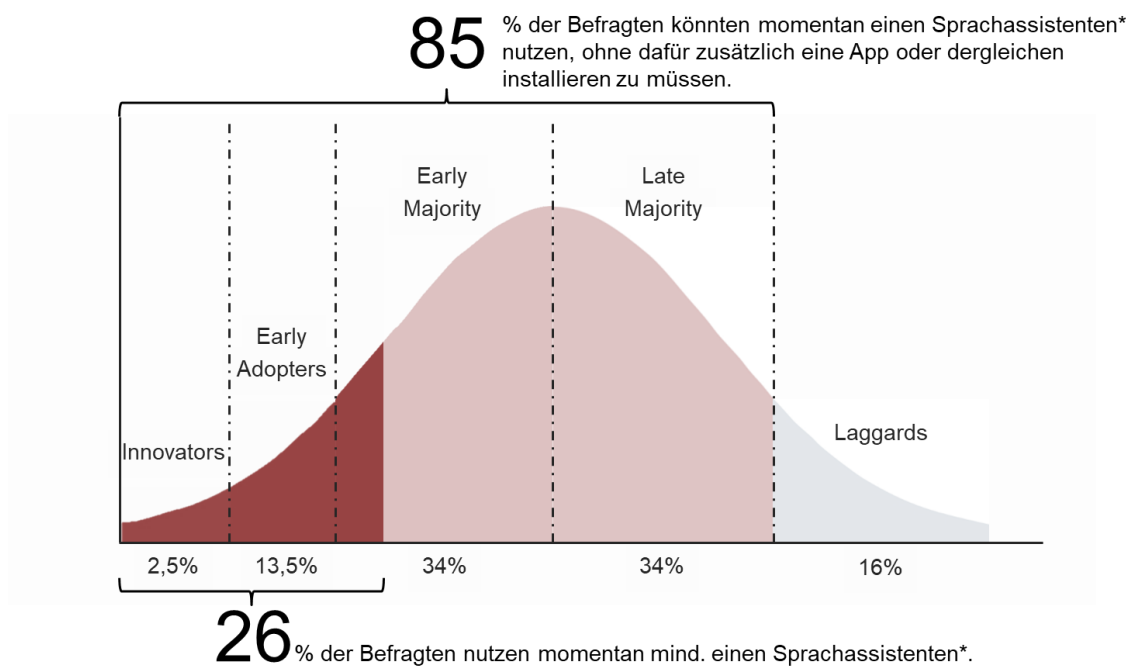


Quelle: Eigene Darstellung basierend auf Daten des GWI (<https://www.globalwebindex.com/> [Letzter Zugriff: 09.04.2019]).

Der Grund dafür, dass Deutschland in der Adoption zurückliegt, liegt nicht so sehr an den Möglichkeiten, die die Konsumenten in Deutschland haben, einen Sprachassistenten zu nutzen. Wie die Ergebnisse der Online-Befragung des WIK bestätigen, haben 85% der Befragten zumindest ein Endgerät, bei dem ein Sprachassistent schon vorinstalliert ist. Somit könnten diese Befragten, ohne weiteren Aufwand zu betreiben, eine neue App oder vergleichbare Anwendungen zu installieren, potenziell sofort mit der Nutzung beginnen. Dennoch sind es lediglich 26% der Befragten, die mindestens einen der populärsten Assistenten<sup>33</sup>, die auf zahlreichen Endgeräten vorinstalliert sind, verwenden. Lediglich 30% nutzen ihre Möglichkeiten somit aus. Abbildung 3-3 verdeutlicht dies visuell anhand der Innovationsadoptionkurve nach Rogers (1962). In dieser Theorie existieren fünf verschiedene Typen von Technologienutzern – auch Adoptoren genannt – die Innovators, die Early Adopters, die Early Majority, die Late Majority und die Laggards. Aktuell befinden sich Sprachassistenten in der dritten Phase der Adoption – der Early Majority.

<sup>33</sup> Betrachtet wurden Amazons Alexa, Apples Siri, Googles Assistant, Microsofts Cortana und Samsungs Bixby.

Abbildung 3-3: Rogers Innovationsadoptionskurve – Sprachassistenten\*



Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=3184. \*Die betrachteten Sprachassistenten sind Alexa, Bixby, Cortana, Google Assistant und Siri.

Obwohl die Adoption von intelligenten Assistenten in Deutschland zunächst sehr langsam zu verlaufen scheint, ähnelt sie jedoch der der anderen Informations- und Kommunikationstechnologien. In 2012 nutzten knapp fünf Jahre nach der Einführung des ersten Smartphones von Apple etwa 36% der Deutschen eines. In 2014 – vier Jahre nach der Einführung des Tablet Computers ebenfalls durch Apple – nutzten 28% einen Tablet Computer (Lutter et al. 2015).<sup>34</sup> Diese Werte liegen in der gleichen Größenordnung wie die der Sprachassistenten. Würde jedoch wie bei den anderen beiden Technologien die Marktdurchdringung anhand der im Umlauf befindlichen Endgeräte, die einen Sprachassistenten vorinstalliert haben, gemessen werden, wäre der Markt in der Logik von Roger schon nahe am Saturationspunkt angekommen.

Wie oben angedeutet, geht in den USA die Adoption von Sprachassistenten deutlich schneller voran als in Deutschland. Laut Taylor et al. (2017) verzeichnete in den USA das Tablet in den ersten fünf Jahren seit Markteinführung eine jährliche Wachstumsrate von 20%, das Smartphone eine von 286% und der intelligente Lautsprecher eine von 653%. Da Sprachassistenten nicht nur in diesen intelligenten Lautsprechern zur Verfügung stehen, sondern ebenso bspw. in Smartphones, Tablets und PCs bzw. Laptops

<sup>34</sup> Es sollte jedoch nicht außer Acht gelassen werden, dass die Nutzung von Smartphones und Tablet Computern mit teilweise hohen Anschaffungskosten verbunden ist. Intelligente Assistenten sind jedoch in der Regel bereits vorinstalliert und können so meist ohne Zusatzkosten genutzt werden.

vorinstalliert sind, unterschätzt diese Wachstumsrate vermutlich noch die tatsächliche Adoption dieser Technologie in den USA.

Der Blick auf die Adoption von intelligenten Lautsprechern ist dennoch sehr informativ, da sie die wesentliche Gruppe von Endgeräten darstellen, bei denen davon ausgegangen werden kann, dass Konsumenten diese mit Blick auf den eingebauten Sprachassistenten kaufen. In der Tat nutzen mit 62% der Besitzer dieser Endgeräte deutlich mehr Konsumenten den eingebauten Sprachassistenten als bei anderen Endgeräten. Derzeit sind insbesondere Alexa, Google Assistant oder Siri über diese Lautsprecher verfügbar. Teilweise unterstützen intelligente Lautsprecher auch mehr als einen Assistenten.

Unabhängig vom Endgerät unterscheidet sich die Adoptionsrate der einzelnen Sprachassistenten deutlich. Der Dienst Cortana, welcher prinzipiell die höchste installierte Basis (theoretische Nutzerbasis) hat, wird von besonders wenigen Konsumenten verwendet. Von einer theoretischen Nutzerbasis von etwa 59% unter den befragten Personen in der Online-Befragung des WIK nutzen nur 7% den Dienst tatsächlich. Den Dienst Alexa hingegen verwenden schon 69% von den 13%, die es prinzipiell tun können.<sup>35</sup> Bisher am wenigsten Anklang findet Bixby. Erst vergangenes Jahr eingeführt, finden sich unter den Befragten lediglich 8%, die den Dienst potenziell nutzen könnten, da sie ein entsprechendes Endgerät besitzen. Jedoch nutzen nur 1% der Befragten Bixby.<sup>36</sup>

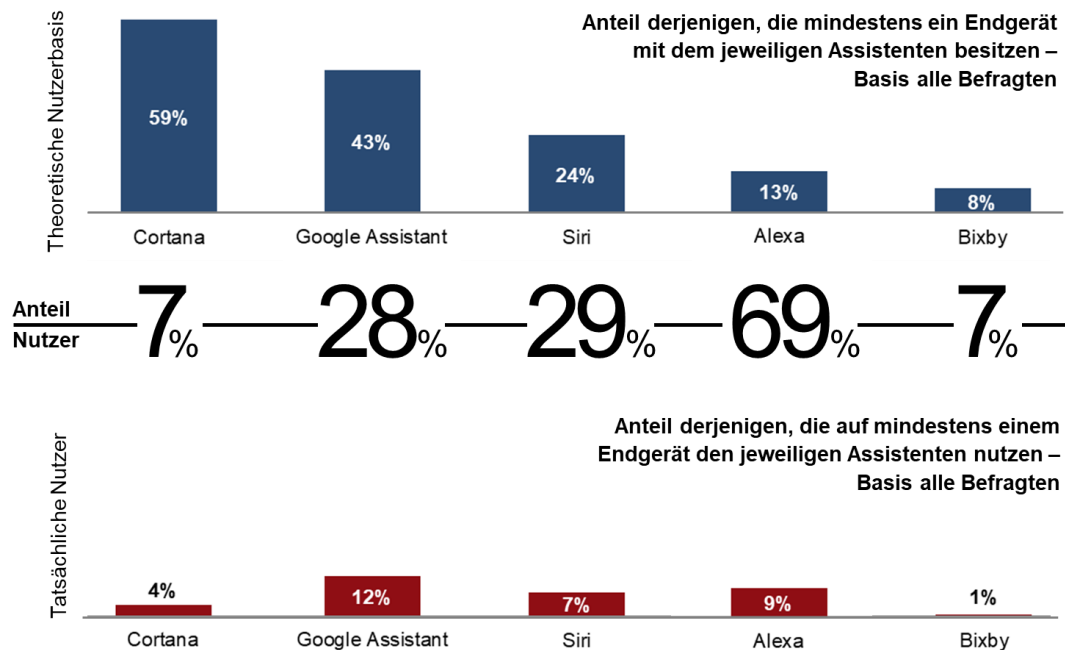
---

**35** Dies korrespondiert mit den hohen Ergebnissen für die Sub-Gruppe der Besitzer von smarten Lautsprechern. Alexa ist der am häufigsten auf diesen Geräten vorinstallierte Sprachassistent.

**36** Es sei an dieser Stelle angemerkt, dass die einzelnen Sprachassistenten, wie Abbildung 3-1 zeigt, zu unterschiedlichen Zeitpunkten in den Markt eingeführt wurden. Dies liefert eine mögliche Erklärung für die unterschiedlichen Anteile der theoretischen Nutzungsmöglichkeit und der tatsächlichen Nutzung.



Abbildung 3-4: Adoption von Sprachassistenten



Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=3184. \*Die betrachteten Sprachassistenten sind Alexa, Bixby, Cortana, Google Assistant und Siri.

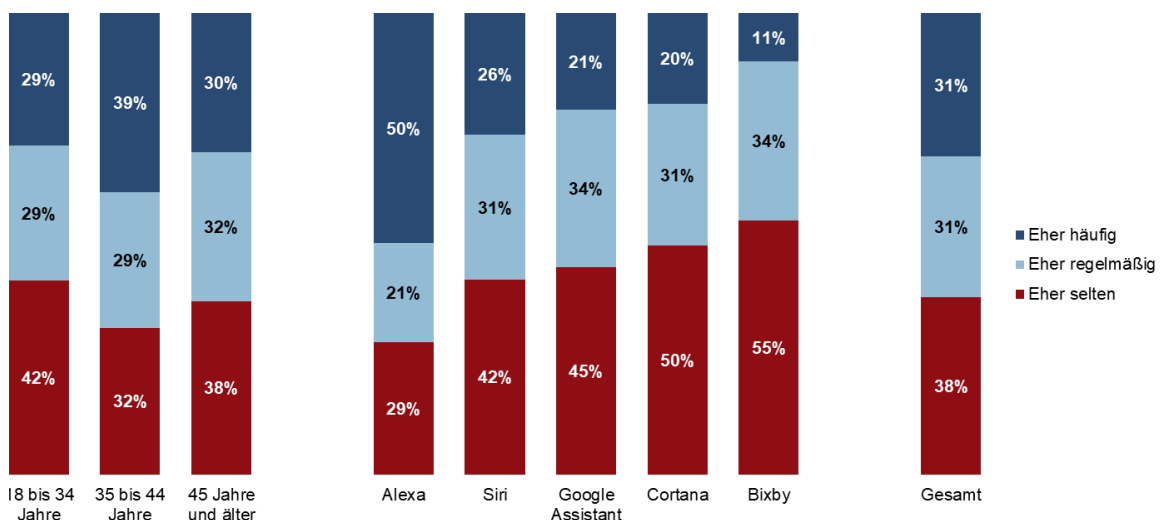
Die allgemeine Nutzungshäufigkeit der betrachteten Sprachassistenten ist in Deutschland in 2018 ebenfalls eher gering. Es können lediglich 18% der Befragten als regelmäßige bis häufige Nutzer klassifiziert werden. Grundsätzlich verteilen sich die Nutzer in etwa gleich – 31% nutzen sie häufig, 31% regelmäßig und 38% selten. Besonders häufig werden Sprachassistenten von den Befragten genutzt, die sich in der Altersgruppe 35 bis 44 Jahre befinden. Nutzer aus der Altersgruppe 18 bis 34 Jahre verwenden intelligente Assistenten hingegen am seltensten. Werden die einzelnen Dienste betrachtet, fällt auf, dass vor allem Alexa häufig von ihren Nutzern verwendet wird. Die anderen Sprachassistenten werden in einem etwa gleichen Umfang von ihren jeweiligen Nutzern verwendet, trotz ihrer unterschiedlichen Nutzeranteile. Wie Abbildung 3-4 zeigt, unterscheiden sich die Nutzerzahlen der Assistenten teilweise sehr deutlich.

Internationale Studien zeigen eine ähnliche Nutzungshäufigkeit unter den Verwendern von Sprachassistenten. Dubiel, Halvey, & Azzopardi (2019) fanden im Rahmen einer Umfrage, dass die meisten Individuen ihre Sprachassistenten einige Male in der Woche nutzen. Darüber hinaus zeigen ihre Daten, dass erfahrene Nutzer den Dienst weniger häufig in Anspruch nehmen.<sup>37</sup> Eine andere Studie von Creative Strategies, die in den

<sup>37</sup> Diejenigen, die einen Sprachassistenten bereits über einen Zeitraum von mehr als 2 Jahren nutzen, verwenden diesen in der Woche weniger häufig als diejenigen, die einen Sprachassistenten weniger als zwei Jahre nutzen.

USA und dem Vereinigten Königreich durchgeführt wurde, belegt ebenfalls, dass trotz der hohen Adoptionsraten in diesen beiden Ländern die tatsächliche Nutzungshäufigkeit noch sehr gering ist. So geben etwa 70% bzw. 60% der Siri- und Google Assistant-Nutzer an, den Service nur manchmal oder gar selten zu verwenden.<sup>38</sup> Auch nach den Befragungsergebnissen von Newman (2018) und Kinsella & Mutchler (2018) sind häufige Nutzer klar in der Minderheit.

Abbildung 3-5: Nutzungshäufigkeit: Sprachassistenten\*



Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=816. Ohne Berücksichtigung derjenigen, die ‚weiß nicht/keine Angabe‘ angegeben haben. Eher selten: Einmal im letzten Monat und einmal in der Woche. Eher regelmäßig: Zwei- bis dreimal in der Woche, Mehr als dreimal in der Woche. Eher häufig: Einmal am Tag, Mehrmals am Tag. \*Die betrachteten Sprachassistenten sind Alexa, Bixby, Cortana, Google Assistant und Siri.

Diese ‚Infrequent User‘ stellen die größte Nutzergruppe von Sprachassistenten dar. Cowan et al. (2017) untersuchte die Nutzung dieser Nutzergruppe mit Blick auf Apples Siri-Sprachassistent in Irland mit Hilfe von Fokusgruppen. Die Autoren fanden, dass der meistgenannte Anwendungsfall innerhalb der Gruppe die Hands-Not-Free-Interaktion war. Das bedeutet, in Situationen, in denen die Anwender ihre Hände nicht frei hatten, wurde Siri als besonders hilfreich erachtet. Bei der Betrachtung der sogenannten ‚Frequent Nutzer‘ von Sprachassistenten fanden Luger & Sellen (2016) heraus, dass unabhängig vom verwendeten Sprachassistenten die interviewten Individuen den Sprachassistenten vor allem dann nutzten, wenn sie viele Aufgaben gleichzeitig zu erledigen hatten. Somit scheint es grundsätzlich die Hands-Not-Free-Interaktion zu sein, die Individuen dazu verleitet und motiviert, Sprachassistenten zu verwenden. Diese Vermutung wird durch eine weitere Studie, die Nutzer unabhängig von ihrer Nutzungsfrequenz in den USA betrachtete, bestätigt. 60% der Nutzer gaben an, besonders die Eigenschaft

<sup>38</sup> <https://creativestrategies.com/voice-assistant-anyone-yes-please-but-not-in-public/> [Letzter Zugriff: 26.06.2019].

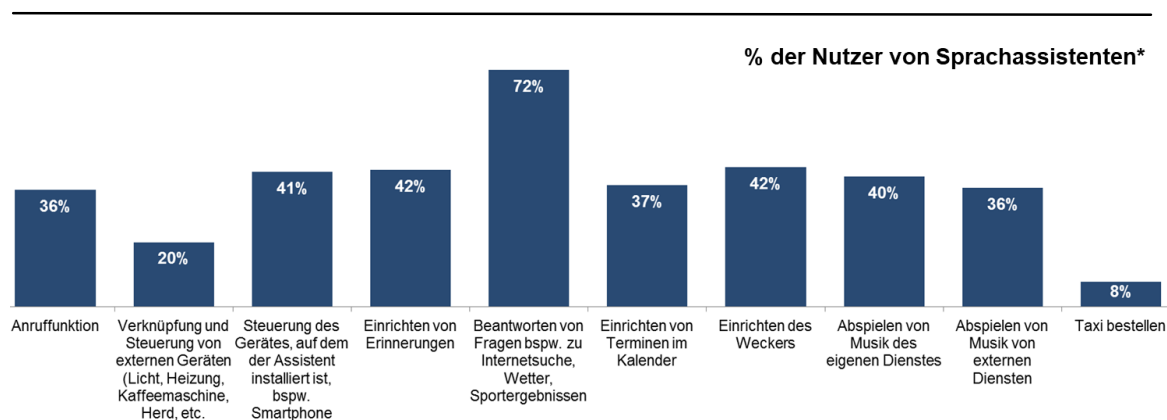
von Sprachassistenten, dass sie auch bedient werden können, wenn die Hände mal nicht frei sind, wertzuschätzen. Daher geben 62% der Nutzer an, die Assistenten zu verwenden, wenn sie Auto fahren (Kinsella & Mutchler 2018). Dies könnte auch begründen, warum die Assistenten vor allem in den stressreichen frühen Morgen- und Abendstunden verwendet werden. Sciuto et al. (2018) konnten beispielsweise anhand von Log-Daten von Alexa-Nutzern zeigen, dass Alexa vor allem in den Morgenstunden um 9:00 Uhr oder in den Abendstunden um 20:00 Uhr vermehrt Verwendung findet. Newman (2018) dokumentiert eine ähnliche Nutzungsverteilung auf den Tag.

## 3.2 Nutzungsmuster

### 3.2.1 Funktionen

Im Rahmen der Online-Befragung des WIK wurden die Befragten ebenfalls nach der Nutzung von zehn für die Assistenten typischen Funktionen befragt (siehe Kapitel 2.2). Alle fünf betrachteten Dienste sind fähig, diese Funktionen auszuführen. Abbildung 3-6 fasst die Verwendung zusammen. Im Durchschnitt nutzen die Nutzer über alle betrachteten Dienste hinweg etwa 3,5 verschiedene Funktionen. Am häufigsten Verwendung findet die Suchfunktion – das Fragen nach Wetter, Sportereignissen oder anderen wenig komplexen Informationen, die zumeist aus dem Internet abgerufen werden. Etwa 72% Nutzer haben 2018 diese Funktion über zumindest einen Sprachassistenten verwendet. Seltener bzw. von weniger Nutzern werden über die Assistenten Taxis bestellt (8%) oder eine Verknüpfung und Steuerung von externen Geräten wie Licht, Heizung oder Herd hergestellt (20%). Die Nutzeranteile für die anderen Funktionen sind hingegen alle ähnlich hoch. Um die 40% der Nutzer nutzen jeweils die Anruf Funktion, steuern das Endgerät, auf dem der Sprachassistent installiert ist, richten sich Erinnerungen, Termine oder einen Wecker ein und spielen Musik ab. Auch hier ist zu erkennen, dass Personen mittleren Alters besonders aktiv sind.

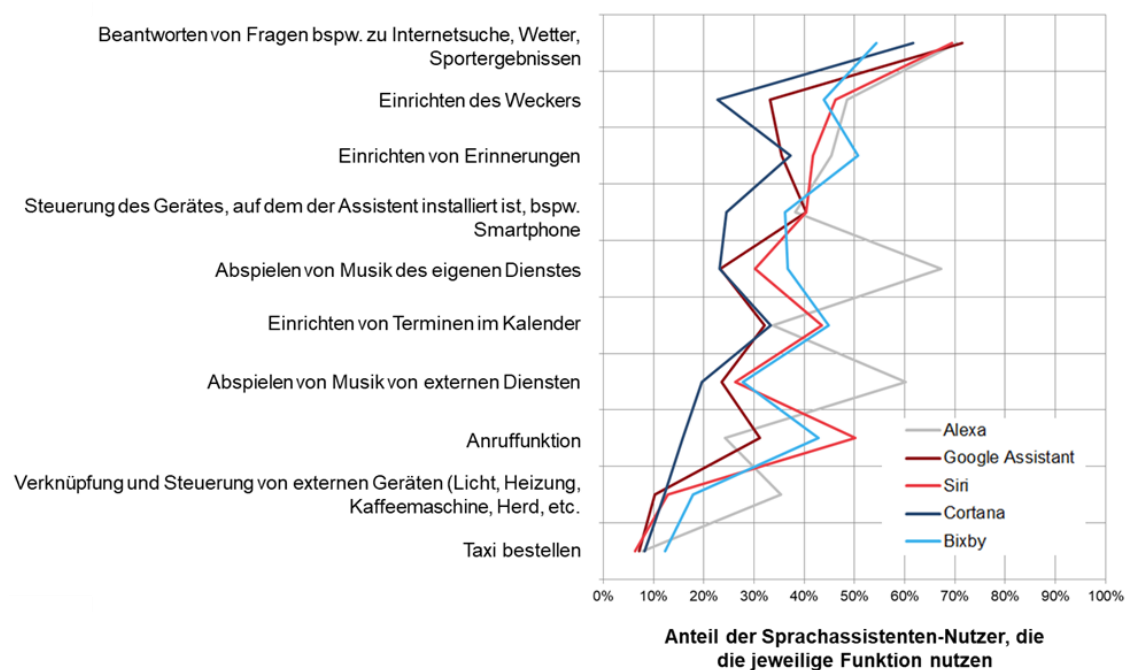
Abbildung 3-6: Genutzte Funktionen – Sprachassistenten\*



Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=816. Ohne Berücksichtigung derjenigen, die ‚weiß nicht/keine Angabe‘ angegeben haben.  
\*Die betrachteten Sprachassistenten sind Alexa, Bixby, Cortana, Google Assistant und Siri.

Obwohl sich die Nutzungsmuster der hier betrachteten Sprachassistenten in Grundzügen ähneln, zeigt Abbildung 3-7 einige bemerkenswerte Unterschiede in der Beliebtheit einzelner Funktionen bei bestimmten Sprachassistenten. Alle Dienste werden von ihren Nutzern für die Suchfunktion verwendet. Der Amazon-eigene Dienst Alexa wird zudem von einem Großteil der Nutzer von Sprachassistenten zum Abspielen von Musik verwendet – sei es über den eigenen Dienst oder über externe Dienste wie Spotify, Deezer oder Tidal. Bixby und Siri werden hingegen verstärkt für die Einrichtung von Terminen, Erinnerungen und Wecker sowie für das Tätigen von Anrufen verwendet. Cortana und der Google Assistant finden insgesamt für wenig Funktionen Gebrauch. Im Durchschnitt wird Cortana für 2,4 Funktionen verwendet und der Google Assistant für etwa 2,9. Die anderen drei Dienste hingegen werden im Schnitt für 3,0 bis 4,2 Funktionen verwendet.<sup>39</sup>

Abbildung 3-7: Genutzte Funktion nach Sprachassistenten



Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=816. Ohne Berücksichtigung derjenigen, die ‚weiß nicht/keine Angabe‘ angegeben haben.

Ein weiterer Unterschied ist, dass die Dienste Alexa, Google Assistant und Siri vor allem von männlichen Nutzern ausgiebig verwendet werden. Diese nutzen die Dienste wesentlich umfangreicher als weibliche Nutzer. Erstere verwenden im Durchschnitt etwa eine Funktion mehr. Bei Cortana und Bixby ist das Verhältnis umgekehrt. Da werden die Dienste umfangreicher von den weiblichen Nutzern verwendet. Aufgrund der relativ geringeren Zahl der Befragten, die Cortana und Bixby verwenden, sollte diese Tendenz jedoch nicht überinterpretiert werden. Die im amerikanischen Raum durchgeführte Un-

<sup>39</sup> Nutzer verwenden etwa 4,2 Funktionen über Alexa, 3,5 Funktionen über Siri und 3,0 Funktionen über Bixby.

tersuchung von PwC (2018) ergab ebenfalls, dass männliche Nutzer eher mit neuen und innovativen Funktionen experimentieren als weibliche.

Bentley et al. (2018) haben einzelne Befehle untersucht. Bei ihrer Betrachtung untersuchten sie knapp 65.000 Interaktionen mit Google Home-Geräten in 88 unterschiedlichen Haushalten in einem Zeitraum von 110 Tagen. Ihre Ergebnisse zeigen, dass vor allem Befehle zu Musik und Informationsbeschaffung gegeben wurden. Die am häufigsten genannten Befehle waren Fragen nach der Uhrzeit, dem Wetter und Anfragen zum Abspielen und Pausieren von Musik oder TV. Lopatovska et al. (2018) dokumentieren, dass zu den beliebtesten Anwendungen von Alexa die Recherche nach Informationen und Entertainmentfunktionen wie das Abspielen von Musik gehören. Eine ähnliche Studie wie Bentley et al. (2018) führten Sciuto et al. (2018) durch. Diese untersuchten 278.654 Befehle, die über unterschiedliche Amazon-Endgeräte an Alexa gestellt wurden. Auch diese Autoren fanden, dass vor allem das Abspielen von Musik eine der am häufigsten verwendeten Funktionen ist – aber auch Fragen zum Wetter und zur Uhrzeit oder das Stellen von Timern spielten eine wichtige Rolle für die Nutzer von Alexa. Anders als bei den oben beschriebenen Studien sehen die Autoren in ihrem Datensatz auch die vermehrte Nutzung von Alexa für Smart-Home-Applikationen, vor allem in den Abendstunden.

Dubiel, Halvey, & Azzopardi (2019) stellten in ihren Befragungen ebenfalls fest, dass die Informationsrecherche, das Abspielen von Videos und Musik sowie die Abfrage des Wetters zu den häufigsten Anwendungsfällen bei der Nutzung von virtuellen persönlichen Assistenten gehören. In weiteren Studien und Befragungen in unterschiedlichen Regionen werden diese Ergebnisse bestätigt. Vor allem die Such- und Entertainment-Funktionen bei Sprachassistenten werden häufig genutzt. Innovative Funktionen wie die Steuerung von externen Endgeräten werden grundsätzlich eher selten genutzt (Purinton et al. 2017, Kinsella & Mutchler 2018, Lopatovska et al. 2018, Newman 2018, PwC 2018, Ammari et al. 2019).

### 3.2.2 Multi-homing

Personen, die mehrere ähnliche Anwendungen verwenden, betreiben sogenanntes Multi-homing.<sup>40</sup> Da die hier betrachteten Sprachassistenten ein Set an gleichartigen Funktionen besitzen<sup>41</sup> und sich die Nutzungsmuster dieser Funktionen, wie oben gezeigt wurde, über die verschiedenen Sprachassistenten hinweg ähneln, kann an dieser Stelle von einander weitgehend gleichwertigen Diensten gesprochen werden. Trotz dieser Ähnlichkeiten scheinen Konsumenten eine recht dominante Präferenz für die Nutzung eines einzelnen Sprachassistenten zu haben. Das zeigen die Ergebnisse der Online-Befragung des WIK. Knapp 78% der Befragten gaben an, nur einen der fünf

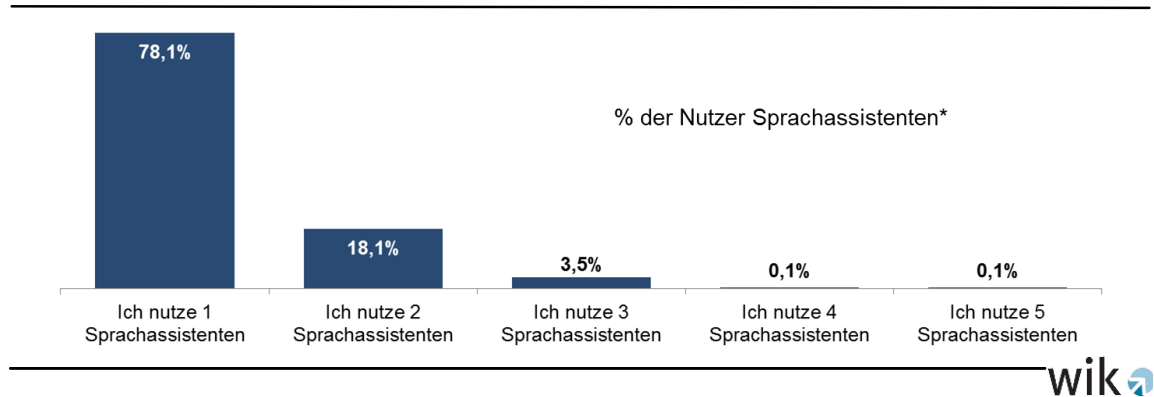
---

<sup>40</sup> Armstrong (2006) beschreibt in einem wettbewerblichen Umfeld Multi-homing wie folgt: „When an agent uses several platforms, she is said to ‚multi-home‘.“ (S.669).

<sup>41</sup> Die existierenden Detailunterschiede in der Umsetzung einzelner Funktionen wurden hier nicht betrachtet. Ebenso wurde außer Acht gelassen, dass es ggf. Qualitätsunterschiede in der Funktionalität einzelner Funktionen der betrachteten fünf Sprachassistenten geben kann.

untersuchten Sprachassistenten zu nutzen.<sup>42</sup> Die Mehrheit der restlichen 22% nutzte zwei Sprachassistenten.

Abbildung 3-8: Muti-homing – Sprachassistenten\*



Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=816. \*Die betrachteten Sprachassistenten sind Alexa, Bixby, Cortana, Google Assistant und Siri.

Wie in Kapitel 3.1 gezeigt wurde, ist vor allem der Google Assistant sehr beliebt, gefolgt von Alexa. Siri, Cortana und Bixby werden hingegen von deutlich weniger Personen verwendet. Somit treten auch am häufigsten Kombinationen auf, in denen einer der beiden erstgenannten Dienste enthalten ist. Denn viele derjenigen, die Siri, Cortana oder Bixby verwenden, nutzen zumeist auch einen der anderen beiden Dienste. Etwa 74% der Nutzer nutzen entweder Alexa oder Google Assistant oder eine Kombination aus Diensten, die diese beiden Assistenten beinhalten.<sup>43</sup> Unter den Nutzern, die mehrere Dienste miteinander kombinieren, sind vor allem die drei Kombinationen Alexa und Google Assistant (23%), Alexa und Siri (19%) sowie Google Assistant und Cortana (17%) sehr beliebt.

Tabelle 3-1: Kombination von Sprachassistenten

Kombination von Sprachassistenten (Basis: Nutzer)					
	Alexa	Google Assistant	Siri	Cortana	Bixby
Alexa	20,5%				
Google Assistant	7,1%	32,1%			
Siri	5,7%	3,9%	19,4%		
Cortana	3,1%	6,1%	1,7%	5,2%	
Bixby	0,6%	1,4%	0,1%	0,9%	1,0%

Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=816. Lesehilfe für Google Assistant und Alexa: 7,1% der Nutzer von Sprachassistenten verwenden eine Kombination von Diensten, die zumindest Google Assistant und Alexa beinhalten.

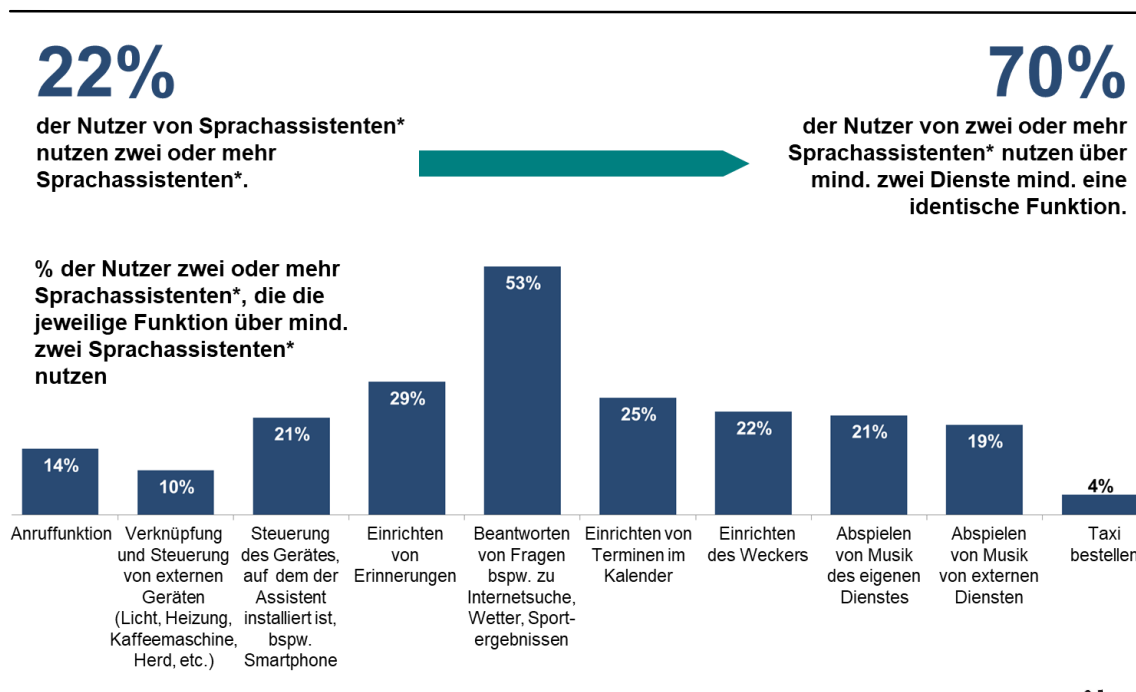
<sup>42</sup> Es ist zu vermuten, dass diese Präferenzstruktur teilweise von der Präferenz für ein bestimmtes mobiles Ökosystem – ANDROID oder iOS – getrieben wird. Obwohl viele der Sprachassistenten ebenfalls unabhängig vom Betriebssystem als mobile App verfügbar sind, kann es sein, dass der Aufwand, diese zu installieren und sie jeweils einzeln aufrufen zu müssen, als zu hoch empfunden wird.

<sup>43</sup> Etwa 26% der Sprachassistenten-Nutzer verwenden keinen der beiden Dienste.

Obwohl 22% der Nutzer mehr als einen Sprachassistenten verwenden, kann hierbei noch nicht direkt von einem grundsätzlichen Multi-homing ausgegangen werden.

Die Analyse ergab, dass von den 22% der Nutzer, die mehr als einen Sprachassistenten nutzen, etwa 70% über mindestens zwei Dienste mindestens eine identische Funktionen verwendeten.<sup>44</sup> Die Nutzung mehrerer Dienste für dieselben Funktionen wird im Folgenden als Funktions-Multi-homing bezeichnet. Abbildung 3-9 zeigt, dass ein großer Anteil derjenigen, die über mindestens zwei Dienste mindestens eine identische Funktion nutzen, vor allem Fragen über diese beantworten (53%). Es sind vor allem diejenigen, die Alexa und Google Assistant (61%) parallel nutzen, die diese Funktion verwenden. Die restlichen Funktionen haben Anteile von je unter 30%.

Abbildung 3-9: Multi-homing: Funktionen



Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018.

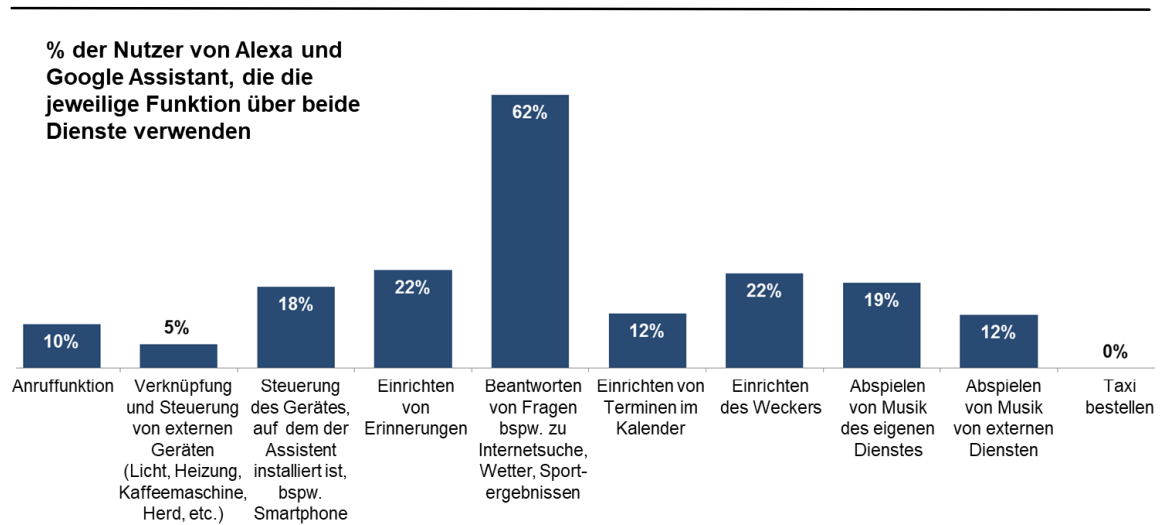
Insgesamt: N=816; Nutzer von zwei oder mehr Sprachassistenten\*: N=180

\*Die betrachteten Sprachassistenten sind Alexa, Bixby, Cortana, Google Assistant und Siri.

Abbildung 3-10 bis Abbildung 3-12 zeigen die Funktionsnutzung der drei beliebtesten Dienstkombinationen. In Abbildung 3-10 und Abbildung 3-12 ist zu erkennen, dass vor allem eine Funktion die gemeinsame Nutzung dominiert: Dies ist jeweils das Suchen nach Informationen bzw. das Beantworten von Fragen. Die anderen Funktionen spielen insgesamt eher eine untergeordnete Rolle. Anders sieht die Verwendung von Alexa und Siri aus. Hier sticht keine der Funktionen dominant hervor. Die Anteile verteilen sich gleichmäßiger über die einzelnen Funktionen.

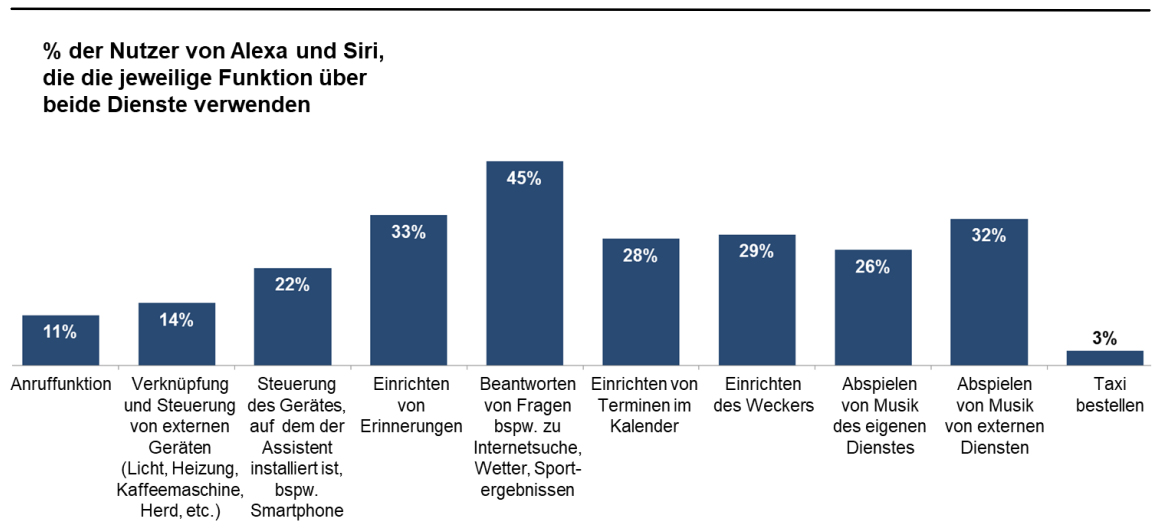
<sup>44</sup> Dies entspricht etwa 6% der gesamten Stichprobe.

Abbildung 3-10: Multi-homing: Funktionen – Alexa und Google Assistant



Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=42.

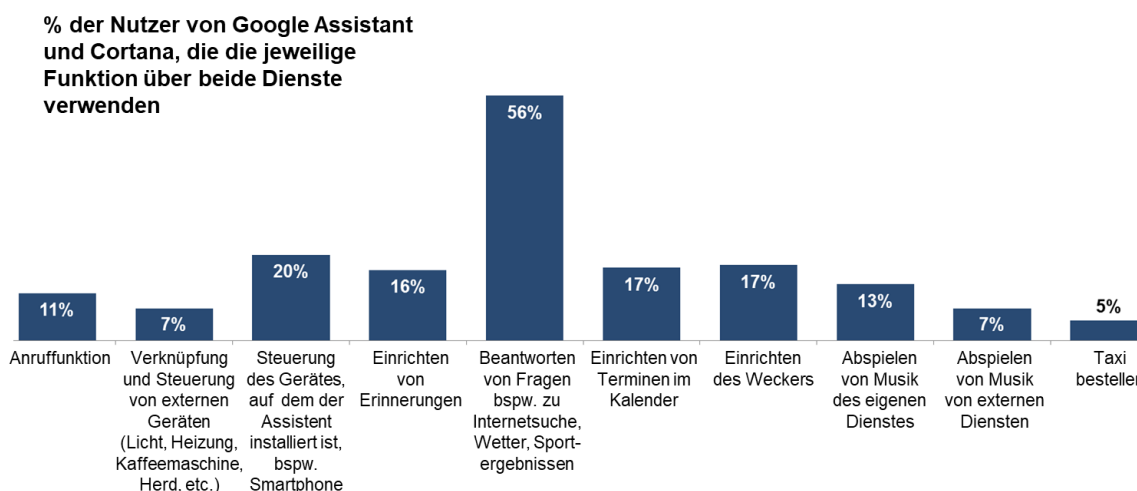
Abbildung 3-11: Multi-homing: Funktionen – Alexa und Siri



Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=34.



Abbildung 3-12: Multi-homing: Funktionen – Google Assistant und Cortana

wik 

Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=32.

Jedoch existiert auch eine zeitliche Komponente im Multi-homing: Werden die Dienste auch im gleichen Umfang genutzt? Wird der Anteil derjenigen, die mindestens zwei Dienste für mindestens eine identische Funktion verwenden, auf die jeweilige Nutzungshäufigkeit heruntergebrochen, ist zu erkennen, dass lediglich knapp die Hälfte (45%) die Dienste faktisch parallel und gleich häufig nutzt.<sup>45</sup> Multi-homing ist bisher noch selten. Knapp 7% aller Nutzer scheinen Multi-homing zu betreiben, wenn sowohl die qualitative als auch die quantitative Ebene des Multi-homing betrachtet wird.

### 3.2.3 Konfiguration

Wie Kapitel 2.3 zeigt haben Konsumenten prinzipiell die Möglichkeit, viele der Standardeinstellungen zu verändern. Insgesamt nutzen jedoch wenige diese Möglichkeit. In der Befragung wurde dabei zwischen der Einstellung der Stimme der Geräte, insbesondere mit Blick auf das Geschlecht der für die Sprachausgabe verwendeten Stimme, und anderen Einstellungen unterschieden. Während die Frage nach der individuellen Einstellung der Stimme von allen Nutzern der Dienste zu beantworten war<sup>46</sup>, gab es für die Frage nach weiteren durchgeführten individuellen Einstellungen nur entsprechende Freifelder, jedoch keine gestützten Abfragen bestimmter Einstellungen.

Nur rund 8% der Nutzer von Sprachassistenten haben das Geschlecht der Stimme ihres Assistenten an ihre persönlichen Präferenzen angepasst. Die absolute Mehrheit belässt die Stimme ihres Assistenten so, wie sie standardmäßig eingestellt wurde.

<sup>45</sup> N=125.

<sup>46</sup> Abgesehen von Alexa bieten alle hier betrachteten Sprachassistenten diese Möglichkeit an. Alexa-Nutzer wurden hier nicht berücksichtigt.

Tabelle 3-2: Änderung von Einstellungen: Stimme

Anteil der Nutzer, die die Default-Stimmen-Einstellung veränderten – Alle Sprachassistenten*							
Geschlecht		Altersgruppen					Gesamt
weiblich	männlich	18 bis 24 Jahre	25 bis 34 Jahre	35 bis 44 Jahre	45 bis 54 Jahre	55 Jahre und mehr	
10,7%	5,7%	8,5%	11,3%	12,5%	10,6%	2,5%	<b>7,9%</b>

Anteil der Nutzer, die die Default-Stimmen-Einstellung veränderten						
		Alexa	Google Assistant	Siri	Cortana	Bixby
Geschlecht	weiblich	Die Stimme ist nicht änderbar	4,9%	22,3%	6,6%	9,0%
	männlich		2,4%	11,2%	2,3%	5,4%
Altersgruppen	18 bis 24 Jahre		4,4%	13,6%	0,0%	8,3%
	25 bis 34 Jahre		4,2%	16,6%	0,0%	9,2%
	35 bis 44 Jahre		4,8%	31,7%	9,6%	9,7%
	45 bis 54 Jahre		8,5%	18,2%	5,1%	8,5%
	55 Jahre und mehr		0,0%	7,4%	2,9%	3,1%
Gesamt		<b>3,6%</b>	<b>16,2%</b>	<b>3,7%</b>	<b>10,5%</b>	

Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=816. Ohne Berücksichtigung derjenigen, die ‚weiß nicht/keine Angabe‘ angegeben haben. \*Die betrachteten Sprachassistenten sind Alexa, Bixby, Cortana, Google Assistant und Siri.

Aber auch bei allen anderen Einstellungen halten sich die Konsumenten zurück. Lediglich 9% der Nutzer haben angegeben, eine weitere Einstellung verändert zu haben. Diese sind namentlich in Abbildung 3-13 erwähnt. Dabei waren die Sprache und die Lautstärke die beiden meistgenannten Einstellungen, die verändert wurden.

Abbildung 3-13: Änderung von Einstellungen

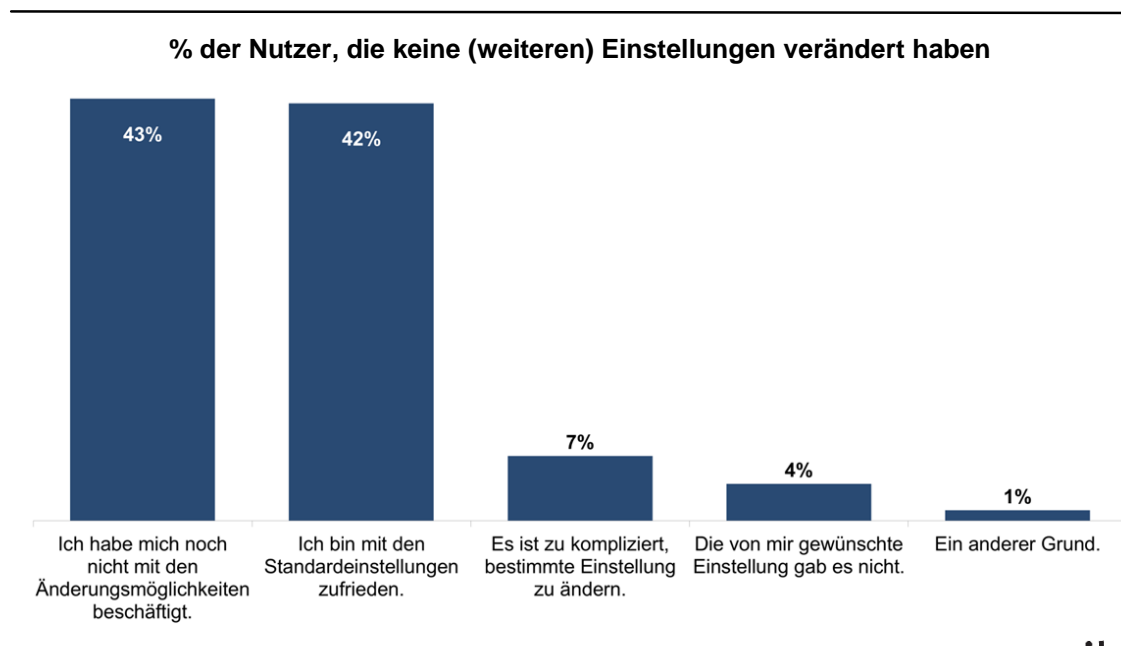


Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=80.

Einstellungen zu Zugriffen auf sensible Daten wie zur Kommunikation, zu Standortdaten und zu Nutzungshistorien werden laut den Angaben der Befragten abgesehen vom Aufmerksamkeitsmodus nicht verändert, obwohl die Befragten an anderen Stellen in der Befragung durchaus ein gewisses Unbehagen und Misstrauen mit Blick auf den Datenschutz der verschiedenen Dienste geäußert haben (siehe Kapitel 4).<sup>47</sup>

Der Hauptgrund, den die Befragten dafür angaben, keine Einstellungen geändert zu haben, ist der Mangel an Interesse. Etwa 43% der befragten Nutzer, die keine Einstellung verändert haben, gaben an, sich nicht mit den Möglichkeiten auseinandergesetzt zu haben. Ungefähr 7% gaben an, es als komplex zu empfinden, Einstellungen zu ändern. Nur 4% gaben an, die Einstellung, die sie ändern wollten, nicht gefunden zu haben. Es sind immerhin 43%, die zufrieden mit den Standardeinstellungen waren.

Abbildung 3-14: Gründe, warum keine (weiteren) Einstellungen der Sprachassistenten\* geändert wurden



Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=736. Ohne Berücksichtigung derjenigen, die ‚weiß nicht/keine Angabe‘ angegeben haben.  
\*Die betrachteten Sprachassistenten sind Alexa, Bixby, Cortana, Google Assistant und Siri.

Ammari et al. (2019) fanden darüber hinaus, dass einigen Konsumenten nicht bewusst ist, wie die Default-Einstellungen angelegt sind. So war vielen der Befragten nicht bewusst, dass ihr Assistent ihre Eingaben in einer Historie abspeichert, die sie einsehen und löschen können. Lau, Zimmerman, & Schaub (2018) führten im gleichen Zusammenhang aus, dass selbst Nutzer, die solche Einstellungsmöglichkeiten kennen, das Öffnen und das Bearbeiten der Log-Dateien zu kompliziert finden.

<sup>47</sup> Dies ist konsistent mit dem so genannten „Privacy Paradox“ (vgl. hierzu Arnold, Waldburger, & Hillebrand (2015) und Barth & de Jong (2017)).

### 3.3 Beziehung zu anderen Diensten und Endgeräten

Im Folgenden werden nun zwei Funktionen genauer betrachtet, die in Bezug auf die mögliche Gatekeeper-Problematik von Sprachassistenten besonders interessant sind, nämlich das Tätigen von Einkäufen sowie der Anschluss externer Geräte. Insgesamt scheinen die Nutzer von Sprachassistenten für komplexere Aufgaben – wie Einkäufen oder Smart-Home-Anwendungen – noch zurückhaltend zu sein.<sup>48</sup>

Die als „Voice Commerce“ oder „Conversational Commerce“ bekannte Art des Einkaufens findet sich aktuell verstärkt in den Medien wieder. Jedoch gibt es nur wenige Studien, die sich den Implikationen und der Reichweite dieser Funktion angenommen haben. Bisher scheint sie nicht nur in Deutschland lediglich rudimentär genutzt zu werden. Im Rahmen einer internationalen Studie haben Buvat et al. (2018) eine Befragungen zu Conversational Commerce in dem Vereinigten Königreich, den USA, Frankreich und Deutschland durchgeführt. In dieser Studie gibt etwa ein Drittel der befragten Nutzer von Sprachassistenten an, schon mal ein Produkt über diesen bestellt zu haben. Doch nicht nur das Einkäufen an sich wird ungern über einen Sprachassistenten durchgeführt, sondern auch alle anderen Serviceleistungen, die mit dem Einkauf verbunden sind, wie beispielsweise der Kontakt zum Kundenservice. In PwC (2018) wurden amerikanische Konsumenten beispielsweise gefragt, ob sie bei bestimmten Situationen eher den traditionellen Weg nehmen oder Sprachassistenten verwenden würden. Obwohl in absoluten Zahlen in den USA öfter Voice Commerce genutzt wird,<sup>49</sup> gaben sowohl beim Kundenservice als auch beim Einkäufen selbst immer jeweils mehr als 60% an, den traditionellen Weg des Anrufens oder des Besuchens des Geschäfts oder Online-Shops zu wählen. Buvat et al. (2018) kamen zu ähnlichen Ergebnissen für die USA, das Vereinigte Königreich, Frankreich und Deutschland. In dieser Studie gab nur eine kleine Gruppe der Befragten (20%) an, lieber den Sprachassistenten zu benutzen als in ein Ladengeschäft oder eine Bankfiliale zu gehen, um einzukaufen oder ihre Geschäfte zu erledigen. Die beiden zuletzt zitierten Studien fanden heraus, dass vor allem die Unsicherheit bezüglich des Bestellprozesses, Datenschutzgründe und ein Mangel an Vertrauen die Konsumenten daran hindern, den Service zu nutzen. Dennoch prognostizieren die Autoren, dass in den nächsten Jahren immer mehr Konsumenten einen Sprachassistenten für Einkäufe verwenden werden.<sup>50</sup>

Faktisch ist Amazons Alexa der einzige intelligente Assistent, über den (in Deutschland) aktuell tatsächlich eingekauft werden kann, da dieser direkt mit dem Amazon Online-

---

<sup>48</sup> Wie oben dargestellt machen einfache Funktionen und Abfragen den Großteil der Nutzung aus. Siehe Abschnitt 3.2.1.

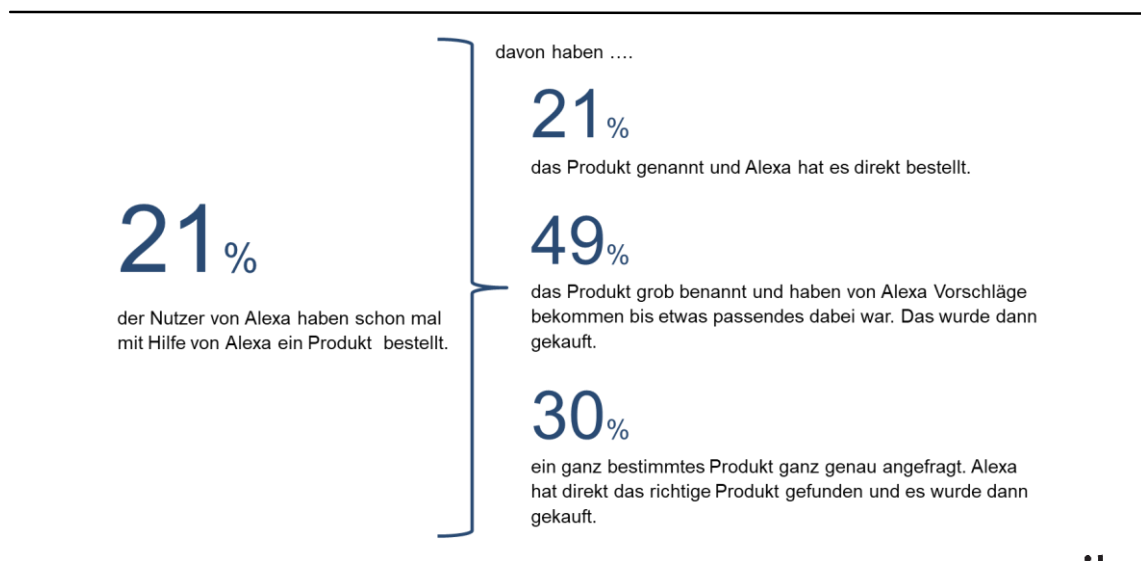
<sup>49</sup> Laut der Studie von Taylor et al. (2017) sollen etwa 35% der Konsumenten schon einmal ein Produkt über einen Sprachassistenten eingekauft haben.

<sup>50</sup> Eine andere Konsumentenbefragung von KPMG und IFH Köln zufolge, hat sogar jeder zweite Nutzer ein Interesse daran, Voice Commerce in Zukunft zu verwenden.  
<https://home.kpmg/de/de/home/newsroom/press-releases/2018/08/voice-commerce-consumer-barometer-2-2018.html> [Letzter Zugriff: 01.07.2019].

Shop verknüpft ist.<sup>51</sup> Obwohl in Deutschland etwa 51% der Nutzer von Alexa theoretisch in der Lage wären, Produkte über Alexa zu bestellen<sup>52</sup>, tun dies lediglich 21%. Somit haben weniger als die Hälfte (41%) der Konsumenten, die theoretisch in der Lage wären, schon einmal eine Bestellung mit Hilfe von Alexa aufgegeben.

Von denjenigen Befragten in der vorliegenden Studie, die über Alexa einkaufen, bestellen die wenigsten ein Produkt direkt und ohne weitere Evaluation über Alexa. Die meisten lassen sich Vorschläge geben und wählen dann ein geeignetes Produkt aus oder haben bereits ein ganz bestimmtes Produkt im Sinn und fragen dieses explizit nach.

Abbildung 3-15: Einkauf über Alexa



Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=272. Ohne Berücksichtigung derjenigen, die ‚weiß nicht/keine Angabe‘ angegeben haben.

Der Anschluss zu externen Endgeräten ist ebenfalls noch nicht weit verbreitet, obwohl diejenigen, die bereits externe Endgeräte angeschlossen haben, mit dem Anschluss, der Konfiguration und der Steuerung zufrieden waren. Alle drei Prozessstufen wurden von den Nutzern zu stets mehr als zwei Drittel bis 90% als einfach empfunden. Das gilt über alle Dienste hinweg.

Betrachtet man die einzelnen Sprachassistenten, so verknüpfen und steuern zwischen 9% (Google Assistant) und 35% (Alexa) den Sprachassistenten mit externen Endgeräten (Abbildung 3-7). In anderen Ländern und vor allem in den USA ist die Verknüpfung von externen Geräten beispielsweise verbreiteter. Laut PwC (2018) verbinden insge-

<sup>51</sup> Hierzu ist anzumerken, dass andere Assistenten wie bspw. der Google Assistant außerhalb von Deutschland durchaus eine Shoppingfunktion anbieten. Im Fall von Google Assistant ist diese an den Dienst Google Shopping gekoppelt.

<sup>52</sup> Um einen Einkauf über Amazon's Alexa tätigen zu können, müssen die Konsumenten sowohl eine Amazon Prime-Mitgliedschaft besitzen als auch die 1-Click-Bestellfunktion nutzen können. In der WIK-Umfrage traf beides bei etwa 51% derjenigen zu, die Alexa verwenden.

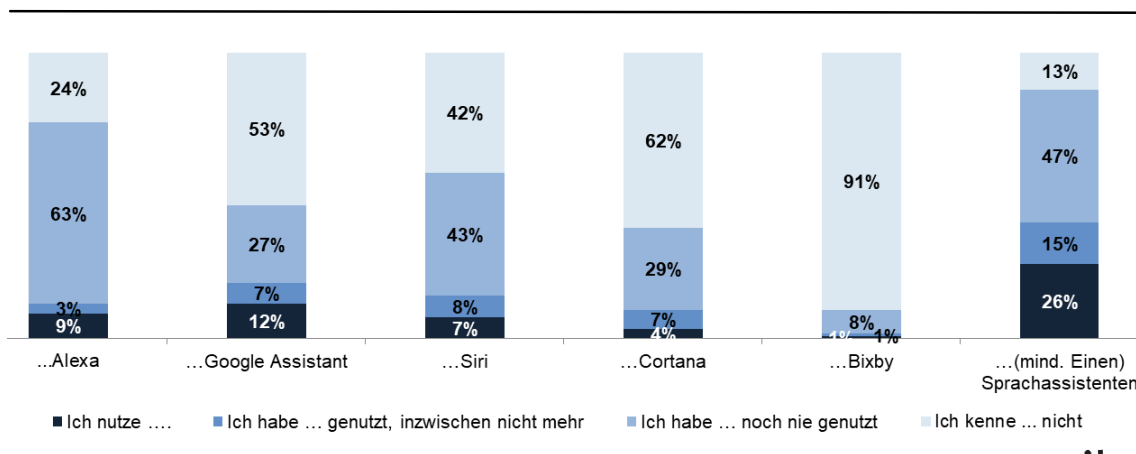
samt 44% der Nutzer ihren Sprachassistenten mit anderen smarten Endgeräten in ihren Haushalten in 2018.

Insgesamt kann festgehalten werden, dass die Nutzung der betrachteten Digitalen Assistenten über deren eigenes geschlossenes Ökosystem hinaus bisher noch geringe Anwendung findet. Dies könnte sich jedoch in den nächsten Jahren zunehmend ändern, da beispielsweise auch das Interesse an Smart Home-Anwendungen steigt (Wagner et al. 2018).

## 4 Einstellung zu Sprachassistenten

Nicht nur die Nutzung an sich, sondern auch die Nutzungshäufigkeit und die Art der Nutzung ist in Deutschland noch nicht sehr stark ausgeprägt. Unter den weniger als 30% der Befragten, die Sprachassistentennutzer sind, nutzt die Mehrheit die Dienste nur selten – also im Durchschnitt etwa einmal pro Woche – und mehrheitlich für das Abfragen von Informationen.

Abbildung 4-1: Nutzung der einzelnen Dienste und Gesamt

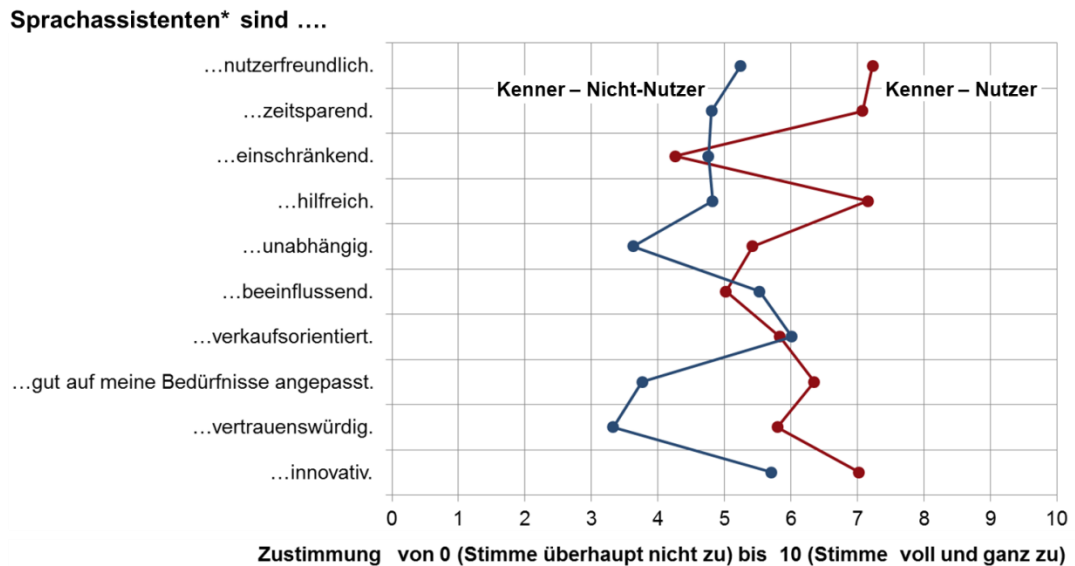


Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=3184. Ohne Berücksichtigung derjenigen, die ‚weiß nicht/keine Angabe‘ angegeben haben.

Um mögliche Gründe für dieses Verhalten zu identifizieren, wurde in der WIK-Befragung die Einstellung der Befragten zu Sprachassistenten im Hinblick auf zehn verschiedene Aussagen untersucht. Dabei teilten sich die einzelnen Aussagen in je fünf negative und positive Einstellungskomponenten. Die Befragten bewerteten die einzelnen Statements anhand einer 11-Punkt-Likert-Skala von „0 – Stimme überhaupt nicht zu“ bis „10 – Stimme voll und ganz zu“. Abbildung 4-2 bis Abbildung 4-4 zeigen die Ergebnisse derjenigen Befragten, die die Sprachassistenten zumindest kennen. Erwartungsgemäß haben die Nutzer einen positiveren Eindruck von den Sprachassistenten als die Nicht-Nutzer (Abbildung 4-2). Die größten Einstellungsunterschiede gibt es im Hinblick auf die Vertrauenswürdigkeit der Dienste und ihre Bedürfnisanpassung. Während die Nicht-Nutzer eher dazu neigen, den Statements nicht zuzustimmen – mit einem durchschnittlichen Zustimmungsscore von 3,5 – würden die Nutzer den Statements eher zustimmen, wenn auch nicht voll und ganz. Dies Ergebnis zeigt sich bei allen fünf betrachteten Sprachassistenten. Darüber hinaus schneidet die Vertrauenswürdigkeit bei allen Kennern durchschnittlich am schlechtesten ab. Insgesamt wird hier nur ein Einstellungsscore von 4,3 erreicht. Auch dieses Ergebnis betrifft alle Assistenten. Hinzu kommt, dass vor allem die Nicht-Nutzer den Aussagen, dass Sprachassistenten verkaufsorientiert, beeinflussend und einschränkend seien, zustimmen. Diese Items erreichen bei den Nicht-Nutzern, wenn auch nicht substantiell, eine höhere Zustimmung als

bei den Nutzern. Auch wenn bei den einzelnen digitalen Assistenten die Einstellungen ähnlich sind, unterscheiden sie sich sehr stark in der Ausprägung (Abbildung 4-3).

Abbildung 4-2: Einstellung zu Sprachassistenten\*: Allgemein – Kenner



Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=2801. Ohne Berücksichtigung derjenigen, die ‚weiß nicht/keine Angabe‘ angegeben haben.  
\*Die betrachteten Sprachassistenten sind Alexa, Bixby, Cortana, Google Assistant und Siri.

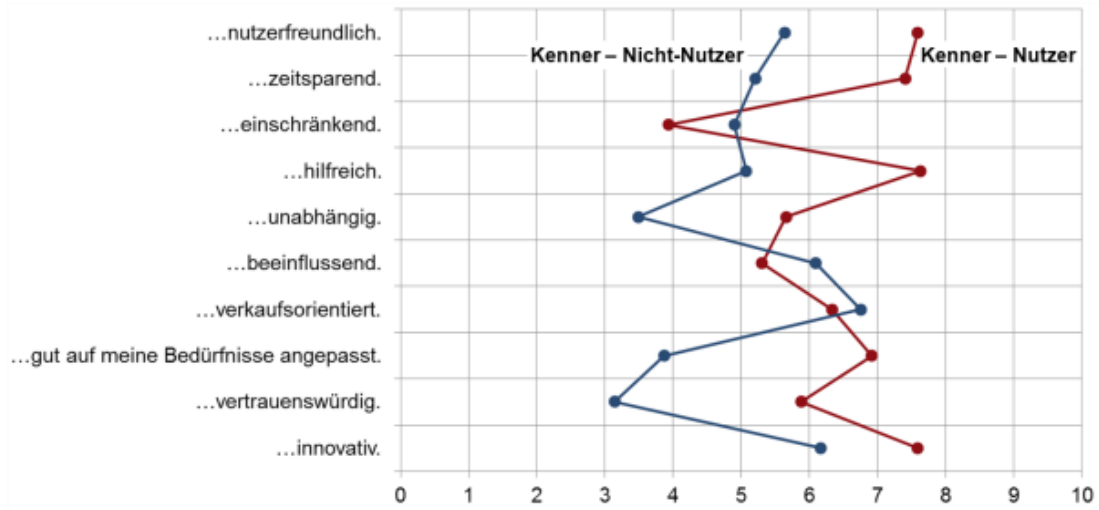
Während die Einstellungen der Alexa-, Google Assistant- und Siri-Kenner nicht wesentlich vom allgemeinen Bild abweichen, ist es bei den Bixby- und Cortana-Kennern etwas anders. Besonders die Cortana-Kenner unterscheiden sich insgesamt stark von den Nutzern in ihren Einstellungen zum Sprachassistenten. Die Bixby-Kenner haben recht ähnliche Einstellungen im Vergleich zu den Nutzern des Dienstes.<sup>53</sup>

<sup>53</sup> Dieses Ergebnis sollte vorsichtig interpretiert werden, da die Stichprobe für Bixby relativ klein ist.

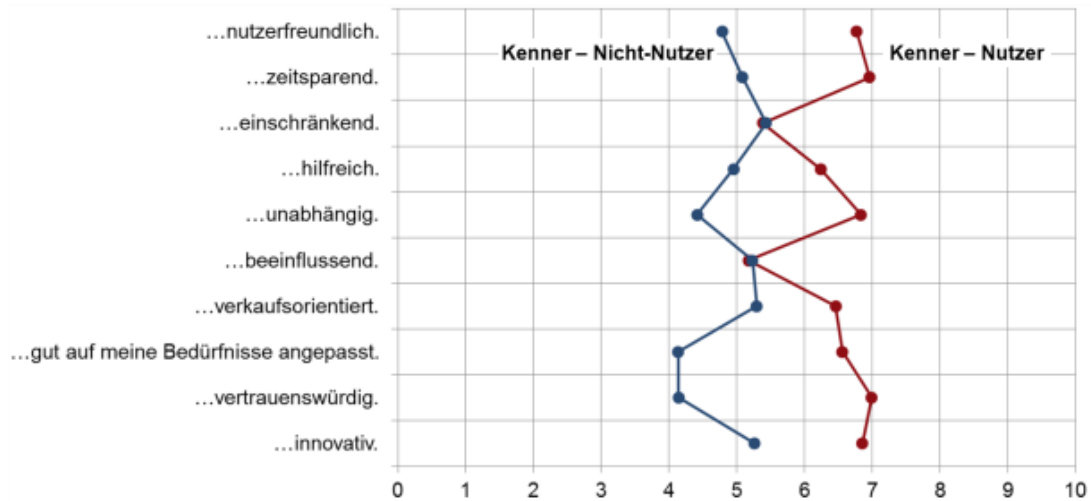


Abbildung 4-3: Einstellung zu Sprachassistenten – Kenner

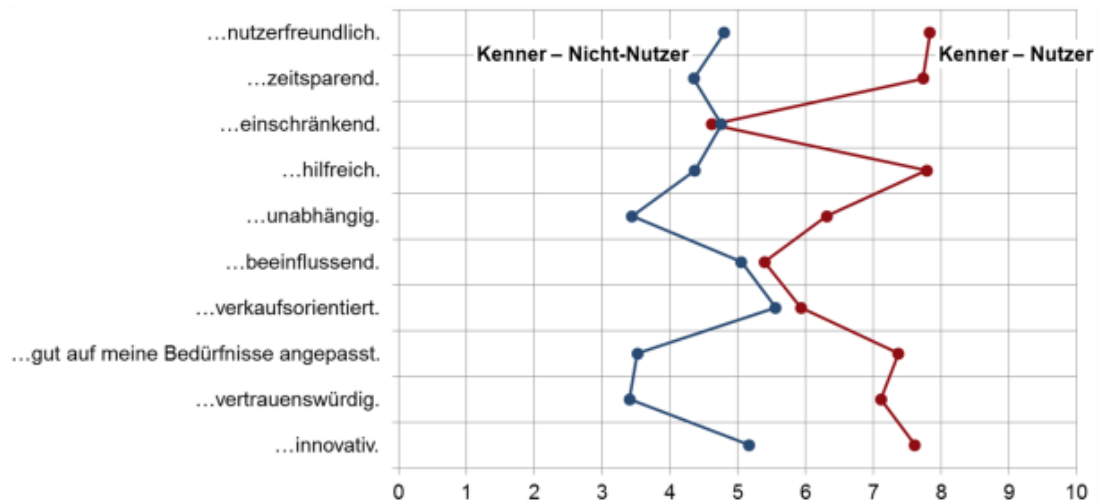
**Alexa ist...**



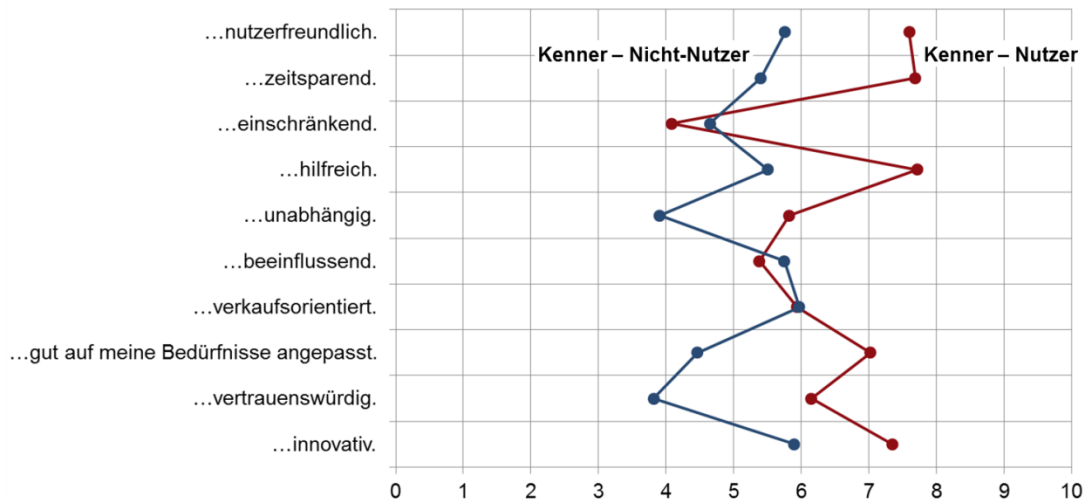
**Bixby ist...**



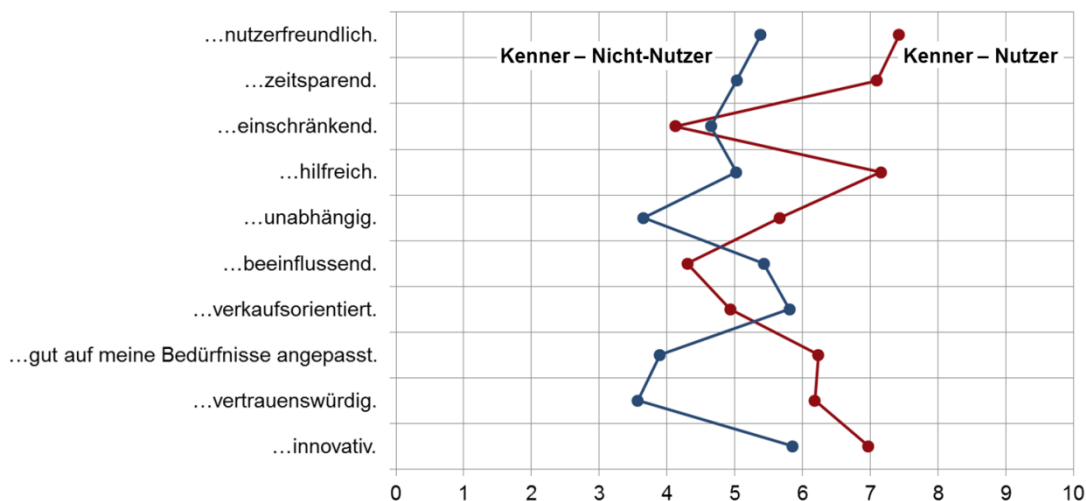
**Cortana ist...**



**Google Assistant ist...**



**Siri ist...**



Zustimmung von 0 (Stimme überhaupt nicht zu) bis 10 (Stimme voll und ganz zu)

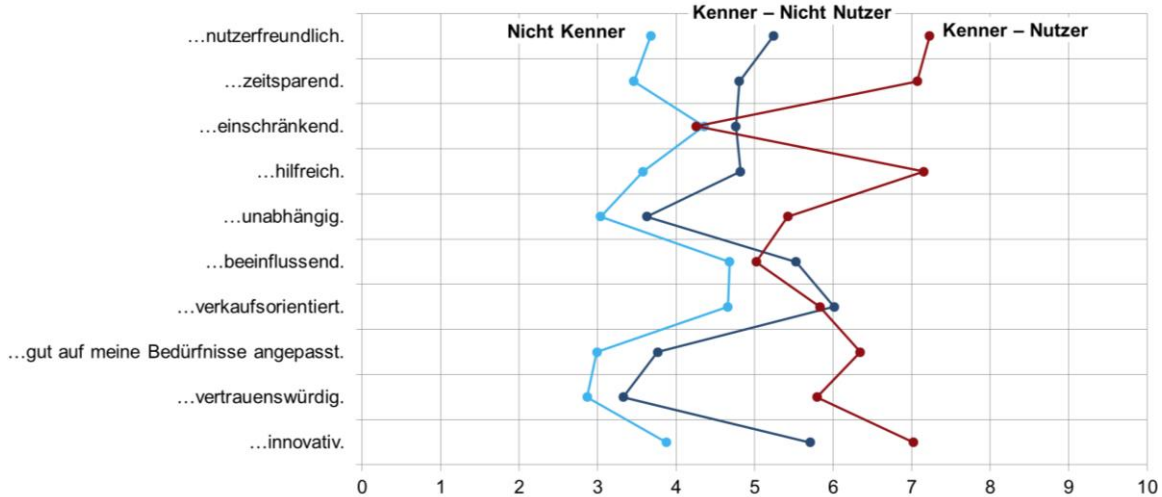
Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. Ohne Berücksichtigung derjenigen, die ‚weiß nicht/keine Angabe‘ angegeben haben. Alexa: N=2412. Bixby: N=307. Cortana: N=1228. Google Assistant: 1489. Siri: N=1842.

Es lässt sich allgemein festhalten, dass diejenigen, die die Assistenten nicht nutzen, grundsätzlich eine negativere Einstellung den Assistenten gegenüber haben als die Nutzer der Dienstes. Dennoch ist ihre Einstellung zu den Assistenten prinzipiell etwas besser als bei denjenigen, die die Assistenten nicht kennen (Abbildung 4-4). Während den negativen Einstellungskomponenten eher zugestimmt wird, wird den positiven eher nicht zugestimmt.

Darüber hinaus zeigen Dubiel, Halvey, & Azzopardi (2019) durch die Untersuchung des Unterschiedes in der Nutzung von virtuellen persönlichen Assistenten zwischen denjenigen Nutzern, die Assistenten häufig nutzen und denjenigen, die sie eher selten ver-

wenden, dass ein positiver Zusammenhang zwischen der Zufriedenheit und der Nutzungshäufigkeit besteht. Die Befragten, die Assistenten häufig nutzen, sind demnach in der Regel zufriedener als diejenigen, die den Dienst selten nutzen.

Abbildung 4-4: Einstellung zu Sprachassistenten\*: Allgemein – Alle

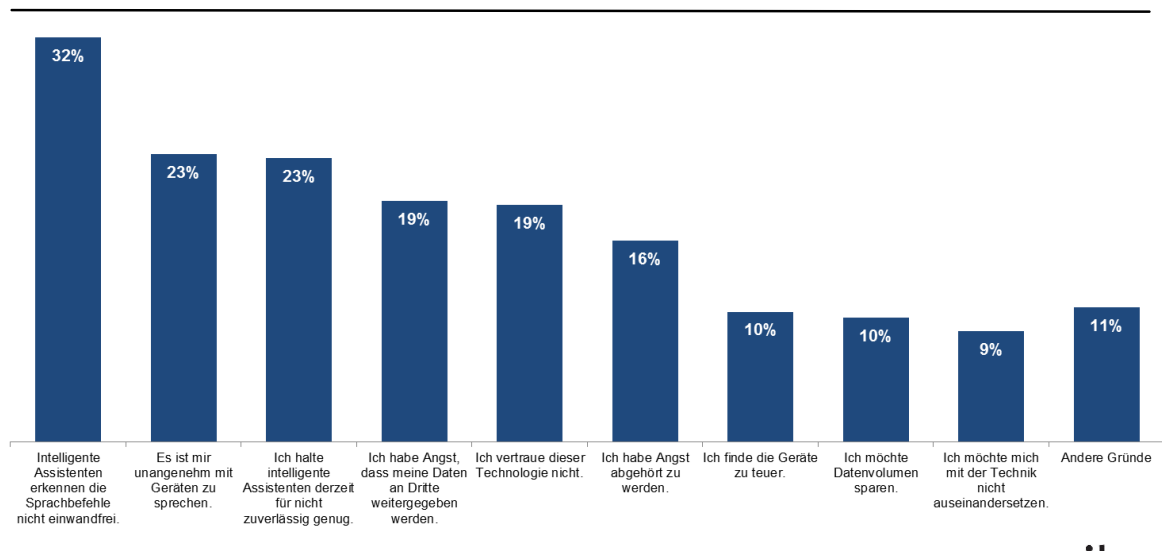


Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. N=3184. Ohne Berücksichtigung derjenigen, die ‚weiß nicht/keine Angabe‘ angegeben haben.  
\*Die betrachteten Sprachassistenten sind Alexa, Bixby, Cortana, Google Assistant und Siri.

Unsere Befragung zeigt, dass viele der Befragten und darunter vor allem Nicht-Nutzer wenig Vertrauen in die Dienste zu haben scheinen. Das passt zu den Angaben der Befragten, die zwar früher zumindest einen Sprachassistenten nutzten, dies aber zur Zeit der Befragung nicht mehr taten. Von diesen stellten viele explizit wegen der Angst, dass vertrauliche Daten erfasst und weitergeleitet werden, die Nutzung der intelligenten Assistenten ein.<sup>54</sup> Genauso wurde die Zuverlässigkeit der Dienste angezweifelt (Abbildung 4-5).

<sup>54</sup> Kürzlich veröffentlichte Presseberichte zur Analyse von aufgezeichneten Audiodaten durch Mitarbeiter verschiedener Hersteller von Sprachassistenten sind vermutlich der Vertrauensbildung in diese Systeme abträglich. Siehe hierzu <https://www.zeit.de/digital/datenschutz/2019-04/amazon-alexa-privatgespraeche-aufzeichnungen-mitschnitte-spracherkennung> [Letzter Zugriff\_ 12.07.2019], <https://onlinemarketing.de/news/amazons-mitarbeiter-hoeren-gespraechе-alexa-an> [Letzter Zugriff: 12.07.2019], <https://www.spiegel.de/netzwelt/gadgets/google-auftragnehmer-hoeren-gespraechе-mit-dem-google-assistant-a-1276894.html> [Letzter Zugriff 12.07.2019]

Abbildung 4-5: Gründe für das Aufhören der Nutzung – ehemalige Nutzer



Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. Ohne Berücksichtigung derjenigen, die ‚weiß nicht/keine Angabe‘ angegeben haben. N=634.

Dieser Mangel an Vertrauen in Sprachassistenten und seine möglichen Auswirkungen auf deren Nutzung wurden bereits in einigen früheren Studien thematisiert. In einer ebenfalls im deutschen Raum durchgeführten Befragung von PwC (2017) gaben mehr als 60% der Befragten an, große Bedenken in Bezug auf die Transparenz für die Anbieter und Möglichkeiten der Manipulation und des Missbrauchs zu haben. In der Umfrage von Buvat et al. (2018) gaben 65% der Nicht-Nutzer an, den Sprachassistenten bei der Sicherheit ihrer Daten nicht zu vertrauen. Auch die Probanden der Fokusgruppen-Untersuchung von Cowan et al. (2017), die sich vor allem aus ‚infrequent‘-Nutzern zusammensetzten, gaben an, dass sie vor allem im Hinblick auf Monetisierungsstrategien des Sprachassistenten Bedenken haben. Unter ihnen herrschte eine grundlegende Ungewissheit darüber, was aufgezeichnet wird und was mit diesen Daten geschieht. In der Umfrage von Buvat et al. (2018) äußerten Konsumenten innerhalb der geführten Interviews auch Unsicherheit bezüglich der tatsächlichen Absichten der Anbieter, die diese Dienste quasi kostenlos anbieten – vor allem in Hinblick auf die Sammlung und den Verkauf von Daten.

Liao et al. (2019) untersuchten die Unterschiede in der Adoption von Sprachassistenten. Sie fanden heraus, dass die Nicht-Nutzer vor allem Bedenken hinsichtlich des Nutzens, den die Assistenten bringen sollen, ihres Designs sowie des Datenschutzes haben. Zusätzlich führten Liao et al. (2019) eine logistische Regression durch, um zu analysieren, welche sozio-ökonomischen und datenschutzrechtlichen Einstellungen und Erfahrungen mit jeweils anderen intelligenten Assistenten dazu beitragen, diese zu nutzen oder nicht. Sie konnten zeigen, dass es einen Zusammenhang zwischen der Nutzung eines Assistenten über ein Smartphone, dem Alter (positiv), dem jeweiligen Grad an digitalen Kompetenzen (positiv), den jeweiligen grundsätzlichen datenschutzrechtli-

chen Bedenken (negativ), dem Vertrauen gegenüber dem Unternehmen, an das die Daten weitergeleitet werden (positiv) und der Tatsache gibt, ob der Nutzer auch Erfahrungen mit einem Assistenten für die Nutzung zu Hause gemacht hat. Bei der Verwendung eines Assistenten, der für die Nutzung im eigenen Heim gedacht ist, existiert ein Zusammenhang mit dem Einkommen (positiv), dem jeweiligen Grad an digitalen Kompetenzen (positiv), den jeweiligen grundsätzlichen datenschutzrechtlichen Bedenken bei Smartphones (positiv), den Bedenken (negativ) und dem Vertrauen (positiv) gegenüber dem Unternehmen, an das die Daten weitergeleitet werden und der Tatsache, ob der Nutzer auch Erfahrungen mit einem Assistenten über ein Smartphone gemacht hat. Das Vertrauen oder die Bedenken wirken sich erheblich auf die Nutzung von Assistenten aus. In einer anderen quantitativen Studie von Han & Yang (2018) konnte ebenfalls ein starker negativer Zusammenhang zwischen Datenschutz- und Sicherheitsbedenken, der daraus resultierenden Unzufriedenheit und dem Zurückschrecken vor der Nutzung von Assistenten festgestellt werden.

Wie oben bereits erwähnt, war ein weiterer negativer Faktor bei den Einstellungen gegenüber den Sprachassistenten und ein wichtiger Grund, Sprachassistenten nicht mehr zu verwenden, dass die Sprachbefehle von den intelligenten Assistenten nicht einwandfrei erkannt wurden. Eine Erklärung hierfür könnten die Studien von Nass & Moon (2000) oder von der Pütten et al. (2010) geben. Diese zeigten und argumentierten, dass Menschen eher dazu geneigt sind, stärker mit Agenten zu kommunizieren, die menschliches Verhalten realistisch simulieren können. von der Pütten et al. (2010) zeigten in einem Experiment, dass je mehr das Verhalten eines digitalen Agenten einem Menschen glich, Probanden desto länger mit diesem kommunizierten. Assistenten, die Befehle nicht verstehen oder nicht einwandfrei darauf reagieren können, nehmen den Nutzern das Gefühl der (simulierten) menschlichen Interaktion. Obwohl die Assistenten kontinuierlich weiterentwickelt werden und diese Verbesserungen bereits für Konsumenten ersichtlich sind, scheinen sie deren Anforderungen an eine echte Konversation immer noch nicht vollumfänglich zu erfüllen.<sup>55</sup>

Das kommt ebenfalls in der bereits in Kapitel 3.1 erwähnten Umfrage von PwC (2018) zum Ausdruck. Dort waren diese Art der Schwierigkeiten der meistgenannte Grund für die Unzufriedenheit der Nutzer mit einem Sprachassistenten. Die Probanden in der Studie von Cowan et al. (2017) gaben sogar an, dass sie sich wünschen, dass Siri mehr mit ihnen interagiert, Rückfragen stellt und Kontexte wahrnimmt, was die Nutzung von Siri hilfreicher und nutzerfreundlicher machen würde. Auch wenn der Dienst von den Studienteilnehmern als „menschenähnlich“ bezeichnet wurde, kamen es ihnen eher vor, roboterähnlich mit ihnen sprechen zu müssen, damit die Befehle verstanden werden. Luger & Sellen (2016) dokumentierten in ihrer Untersuchung, dass Individuen, die einen Sprachassistenten verwenden, meist damit ansetzen, auf eine natürliche Art und Weise mit dem Assistenten zu interagieren, jedoch die Spracheingaben simplifizieren müssen, bis das gewünschte Ergebnis erreicht wird. Jiang, Jeng, & He (2013) untersuchten in einem Experiment, wie die Probanden bei einem Auftreten eines Verständ-

---

<sup>55</sup> Vgl. ebenfalls Porcheron et al. (2018).

nisfehlers seitens des Assistenten die Spracheingabe ändern. Die meisten gingen neben der Änderung der Reihenfolge der Wörter dazu über, einzelne Wörter aus der Eingabe herauszunehmen. Laut den Autoren führen solche Fehler zu negativen Gefühlen bei den Konsumenten.

Darüber hinaus haben Kopp et al. (2005) und Benbasat & Wang (2005) beispielsweise gezeigt, dass je menschenähnlicher und natürlicher die Kommunikation mit derartigen Agenten ist, desto eher wird Vertrauen aufgebaut und desto öfter werden die Dienste verwendet.<sup>56</sup> Einen differenzierteren Einblick hierüber geben beispielsweise Lopatovska et al. (2018). Die Autoren führten eine Befragung unter Alexa-Nutzern durch. Ihre Ergebnisse zeigten eine grundsätzliche Zufriedenheit mit der Performance von Alexa, wenn es vor allem um die einfachen und integrierten Funktionen wie die Wetterabfrage, das Abspielen von Musik, das Einstellen des Weckers usw. ging. Unmut herrschte jedoch bei der Recherche nach Informationen – also bei etwas komplizierteren Aufgaben. Ähnlich waren auch die Ergebnisse von Kiseleva et al. (2016) bei der experimentellen Untersuchung der Zufriedenheit bei der Nutzung von Cortana unter Mitarbeitern eines IT-Unternehmens. Die Autoren fanden ebenfalls einen negativen Zusammenhang zwischen der Zufriedenheit der Nutzer und der Komplexität der an Cortana gestellten Aufgabe. So sahen sie, dass die Probanden zufriedener waren, wenn das Gerät, auf welchem Cortana installiert war, selbst durch Cortana gesteuert wurde und weniger zufrieden, wenn über Cortana nach Informationen gesucht wurde. Je komplexer die gewünschte Interaktion mit Cortana wurde, desto weniger zufrieden waren die Probanden, da Cortana den Anforderungen nicht genügte. Entsprechend fanden Hashemi et al. (2018) heraus, dass vor allem die Fähigkeit eines Assistenten, mehrere aufeinanderfolgende Anfragen zu bearbeiten, die Nutzer zufrieden stellt. Die Probanden des Experiments von Hashemi et al. (2018) waren grundsätzlich knapp 30% zufriedener, wenn der Sprachassistent in der Lage war, vorangegangene Dialoge zu ‚merken‘ und dadurch Folgeanfragen schneller und effektiver zu bearbeiten als ein System, welches dies nicht konnte. Im letzteren Fall verliert der Assistent nach jeder Anfrage den Kontext, was die Bedienung des Assistenten schwieriger und langatmiger macht (Kiseleva et al. 2016). Luger & Sellen (2016) hoben zusätzlich die geringe Frustrationsgrenze der Nutzer hervor. Schließen die Endgeräte nach mehrfacher Befehlseingabe die Aktion nicht wunschgerecht ab, wird der Dienst meist nicht mehr für diese Aktion genutzt oder grundsätzlich weniger.

Einen weiteren wichtigen Einblick in diese Thematik bieten die Studien Buvat et al. (2018) und Purington et al. (2017). Buvat et al. (2018) fragten ihre Probanden, welche Eigenschaften ein Sprachassistent besitzen soll, damit sie bereit sind, diesen zu nutzen. Mehrheitlich wurde hier angegeben, dass vor allem die Fähigkeit des Assistenten, sie zu verstehen und eine Konversation mit ihnen zu führen, dazu beitragen würde, von dem Dienst überzeugt zu sein. Purington et al. (2017) analysierten hingegen den Einfluss des Grades der Personalisierung des Assistenten Alexa unter den Nutzern und die Zufriedenheit mit dem Service, den Alexa bietet, anhand der Auswertung von Rezensi-

---

<sup>56</sup> Zitiert nach Sciuto et al. (2018).

onen auf Amazon.com. Zunächst fanden sie ähnlich wie bereits die oben beschriebenen Studien, dass die Zufriedenheit mit dem Sprachassistent bei einer zunehmenden Anzahl von technischen Problemen abnimmt. Sie fanden aber auch heraus, dass die Zufriedenheit stark mit der Personifizierung von Alexa zusammenhängt. Je mehr ein Nutzer Alexa als tatsächliche Person wahrnimmt, desto zufriedener ist er mit dem Dienst, selbst wenn dieser nicht einwandfrei funktioniert. Jedoch war der Grad der Personifizierung unter den Befragten sehr gering; die meisten nahmen Alexa nicht als Person wahr.

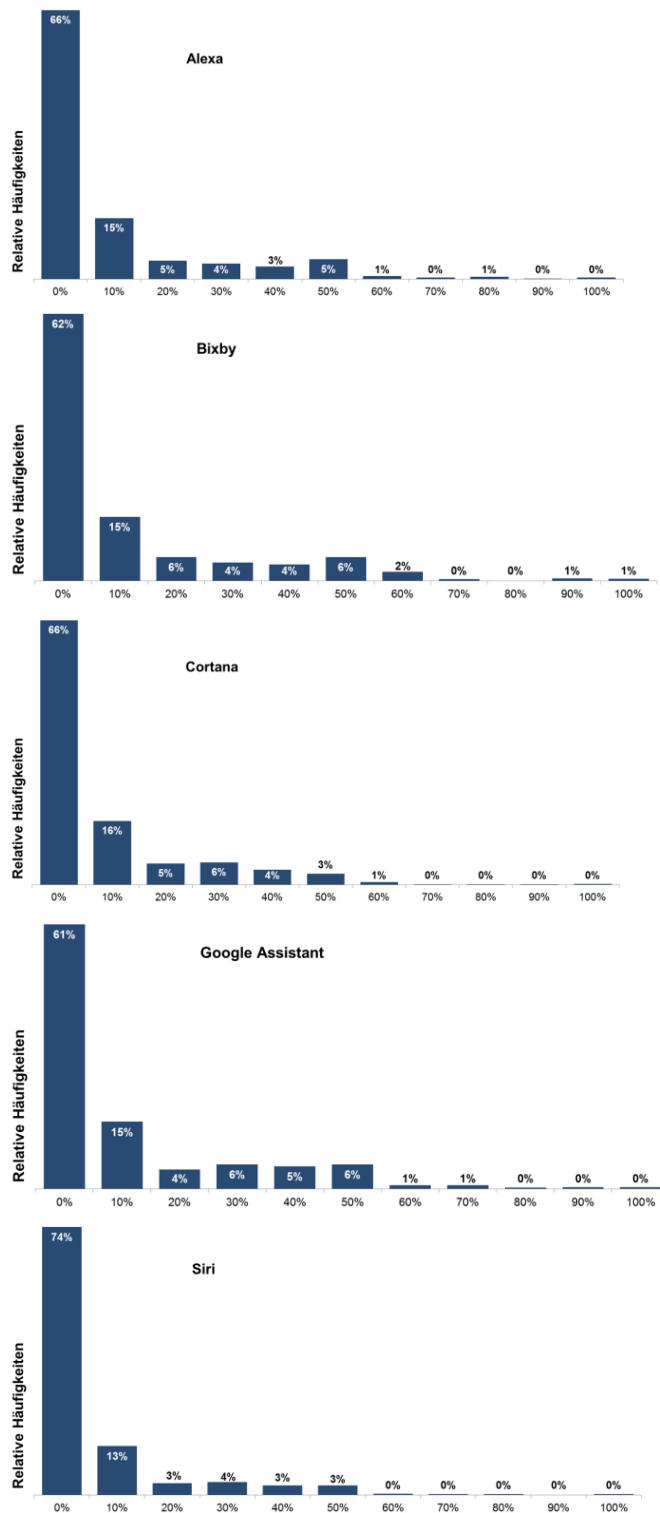
Han & Yang (2018) untersuchten in diesem Zusammenhang die These, dass durch den menschlich-ähnlichen Charakter von Sprachassistenten diese nicht nur personifiziert werden, sondern dass die Nutzer eine soziale Beziehung mit ihnen eingehen und dies die Zufriedenheit bestimmt. Die Autoren bestätigten, dass für den Erfolg eines Assistenten dieser noch menschlichere Formen annehmen müsste, um eine soziale Beziehung zu den Assistenten aufzubauen und zu verstärken, was sich letztlich positiv auf die Zufriedenheit und die zukünftige Nutzung auswirken würde.

Während die bisher erwähnten Studien allgemein den Aspekt der menschenähnlichen Kommunikation von digitalen Assistenten und ihrer Auswirkungen auf die Nutzung betrachten, führten Clark et al. (2019) eine wesentlich differenziertere Analyse durch. Im Rahmen von Interviews wurde untersucht, welche Charakteristika in Konversationen tatsächlich als wichtig erachtet werden und wie sich die Einstellung diesen gegenüber verändert, wenn es um die Interaktion mit digitalen Agenten geht. Die Autoren fanden, dass vor allem Attribute wie ein gegenseitiges Verständnis, Vertrauenswürdigkeit und das aktive Zuhören des Gegenübers wichtig für Individuen bei einer zwischenmenschlichen Konversation sind. Dieselben Faktoren spielten für die Befragten auch bei der Kommunikation mit digitalen Agenten eine wichtige Rolle. Jedoch ließ sich aus den Antworten ableiten, dass aktuelle digitale Assistenten nicht in der Lage sind, den Anforderungen in gleichem Maße zu genügen wie Menschen. Diese Aspekte seien bei Agenten sehr funktional und einseitig ausgelegt.

Zusammenfassend zeigen diese Studien ähnlich der Argumentation von Porcheron et al. (2018), auf die in Kapitel 3 eingegangen wurde, dass die aktuellen Sprachassistenten noch weit davon entfernt sind, als echte „conversational agents“ im Alltag der Konsumenten integriert zu sein.

Vielleicht liegt es an diesen Gründen, dass viele derjenigen, die die jeweiligen Dienste noch nie verwendet haben, auch nicht in Betracht gezogen haben, diese zumindest auszuprobieren. Lediglich zwischen 10% (Bixby) und 18% (Cortana) haben den Dienst zumindest schon mal ausprobiert. Der Anteil von Cortana ist demnach am höchsten, die restlichen bewegen sich zwischen 10% und 13%. Diese Nicht-Nutzer geben an, dass es sehr unwahrscheinlich ist, dass sie die Sprachassistenten in naher Zukunft nutzen werden (Abbildung 4-6).

Abbildung 4-6: Zukünftige Nutzungsintention: Kenner aber Nicht-Nutzer (Noch Nie)



Nutzungswahrscheinlichkeit – 0% (so gut wie ausgeschlossen, auf keinen Fall) bis 100% (sicher, praktisch sicher)

Quelle: Eigene Darstellung; Daten stammen aus der Online-Umfrage des WIK aus dem Jahr 2018. Ohne Berücksichtigung derjenigen, die ‚weiß nicht/keine Angabe‘ angegeben haben. Alexa: N=2009. Bixby: N=247. Cortana: N=892. Google Assistant: 855. Siri: N=1336.



## 5 Herausforderung für Wettbewerb und Regulierung

Um mögliche Herausforderungen für Wettbewerb und Regulierung durch eine breite Adoption und häufige Nutzung von Sprachassistenten einschätzen zu können, ist es zunächst wichtig zu erkennen, dass es bei dieser Technologie eben nicht nur um eine neue Mensch-Maschine-Schnittstelle oder Benutzeroberfläche geht. Schon der kurze Überblick zur Evolution, zu den Bausteinen und den Funktionen von Sprachassistenten zeigte deutlich, dass Sprachassistenten wesentlich mehr leisten, als die Eingabe von Informationen und Befehlen von der Tastatur und Computermaus auf die Ebene des gesprochenen Worts zu verlagern. Typischerweise sind diese Systeme mit internen und externen Datenquellen verbunden, um Fragen der Nutzer zu beantworten bzw. nachgefragte Aktionen auszuführen.

Insofern spielen Netzwerk- und Feedbackeffekte eine besondere Rolle für Sprachassistenten. Abbildung 5-1 gibt einen Überblick über die Wirkung dieser Effekte. Die Implementierung von Sprachassistenten in immer mehr und unterschiedliche Endgeräte ermöglicht starke Lernkurveneffekte und macht es aus ökonomischer Perspektive effizient, wenn möglichst viele Nutzer einen bestimmten Sprachassistenten nutzen.<sup>57</sup> Dabei gilt es zwei wesentliche Arten von positiven Feedbackschleifen zu unterscheiden:

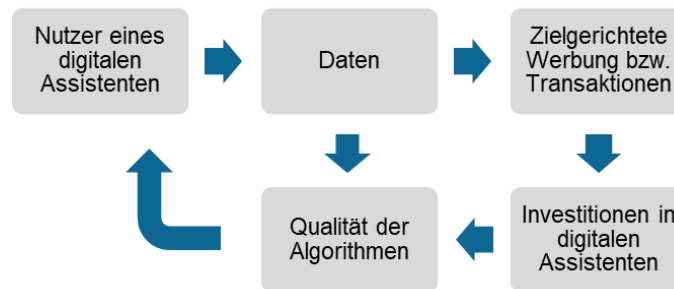
- (1) bei der Nutzer/Endgeräte-Feedbackschleife ist ein Sprachassistent mit einer großen Nutzer- bzw. Endgerätebasis in der Lage, noch mehr Daten zu sammeln, um die zugrundeliegenden Algorithmen weiterzuentwickeln oder die Entwicklung neuer und besserer Skills anzustoßen, die die Qualität und den Funktionsumfang des Sprachassistenten verbessern und somit neue Nutzer bzw. Endgerätehersteller anziehen. Dies geschieht durch die Öffnung des Dienstes für Dritte, um möglichst viele Arten von Endgeräten zu integrieren und schließlich innerhalb eines Ökosystems kontrollieren zu können.
- (2) bei der Monetarisierungs-Feedbackschleife analysieren Sprachassistenten die gesammelten Nutzer- bzw. Endgerätedaten einschließlich ihrer Metadaten (welche durch Sprachschnittstellen im Vergleich zu etablierten Systemen aufgrund der vielfältigen Informationen, die Sprache, Tonalität, usw. über die Nutzer bieten, massiv zunehmen werden)<sup>58</sup> und verwenden die gewonnenen Erkenntnisse bspw. für zielgerichtete Werbung bzw. Transaktionen, um den Sprachdienst zu monetarisieren. Durch diesen Prozess stehen neue Mittel für Investitionen zur Verfügung, die eine Verbesserung der Qualität des Sprachassistenten bzw. der Algorithmen oder der technischen Infrastruktur ermöglichen, welche wiederum noch mehr datengenerierende Nutzer bzw. Endgerätehersteller anzieht.

---

<sup>57</sup> Aufgrund der hohen Kontextabhängigkeit von Spoken-Dialogue-Systemen, die die Grundlage für Sprachassistenten bzw. deren Actions, Apps oder Sills bilden, muss dies nicht unbedingt der Fall sein. Es ist denkbar, dass für bestimmte Zwecke effiziente Module zu komplexeren Systemen integriert werden, wobei jedes der Module von unterschiedlichen Anbietern stammt. Dabei kann eine einheitliche oder uneinheitliche technologische Basis zugrunde liegen.

<sup>58</sup> Darüber hinaus bestehen zusätzliche Möglichkeiten, Metadaten über die Umgebung mit Hilfe zusätzlicher Sensoren zu sammeln. Dazu können bspw. die Anzahl, Position, Herzfrequenz, usw. der in der Nähe befindlichen Personen gehören.

Abbildung 5-1: Netzwerk – und Feedbackeffekte von Sprachassistenten



Quelle: WIK.

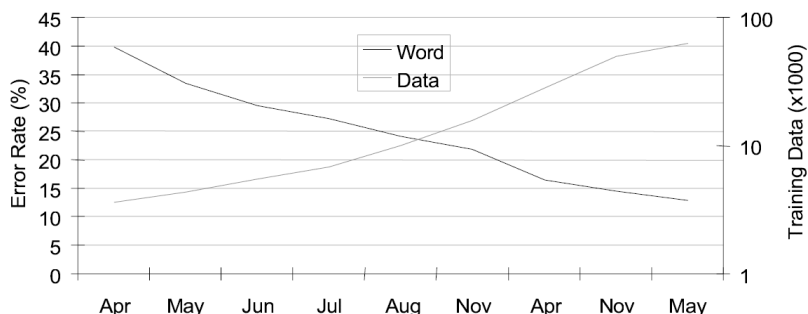
wik

Zue & Glass (2000) zeigen deutlich, dass es sich gerade bei Sprachassistenten um eine Technologie handelt, bei der es darauf ankommt, Zugang zu möglichst vielen Nutzern von solchen Systemen zu haben (vgl. auch Abbildung 5-2), um eine hohe Qualität des Systems zu gewährleisten. Deshalb haben die meisten der frühen Systeme als Telefondienste zur Abfrage von Informationen wie Fahrplaninformationen oder Wetterberichte gestartet, um möglichst viele potenzielle Nutzer zu gewinnen, die die Systeme mit echten Fragen und somit Daten füttern. Die Autoren zeigen deutlich, dass gerade in der Anfangsphase zusätzliche Daten einen exponentiellen Vorteil für die Verbesserung des Spracherkennungssystems ergeben (vgl. Abbildung 5-2).

Heute gibt es zahlreiche frei verfügbare Korpusse an gesprochener Sprache, die zum Trainieren von Spoken-Dialogue-Systemen verwendet werden können (Serban et al. 2018). Jedoch ist anzumerken, dass die Verwendbarkeit von verfügbaren Korpusen zum Training von Spoken-Dialogue-Systemen je nach Anwendungszweck eingeschränkt sein kann, da die Trainingsdaten im Idealfall möglichst nahe am Kontext der geplanten Verwendung bspw. der zu entwickelnden Skills sein sollten, um möglichst gute Ergebnisse zu erzielen (McTear, Callejas, & Griol 2016). Weiterhin kann ein limitierender Faktor für die Qualität des Ergebnisses sein, dass die Daten in uneinheitlicher Aufnahmequalität vorliegen (Wang, Takaki, & Yamagishi 2016). Das ist insbesondere dann der Fall, wenn die Aufnahmequalität nicht den spezifischen technischen Eigenschaften des Zielsystems entspricht.

Neben der Datengrundlage spielen die verwendeten Methoden aber eine mindestens ebenso zentrale Rolle für die Qualität des zu entwickelnden Dienstes. Wang, Takaki, & Yamagishi (2016) zeigen bspw. deutliche Unterschiede in der Ergebnisqualität verschiedener neuronaler Netze bei exakt gleichen Inputdaten.

Abbildung 5-2: Vergleich der Spracherkennungsfehlerrate mit der Anzahl der vorhandenen Sprachinteraktionen mit Nutzern



Quelle: Zue & Glass (2000) S. 1174.

Monopolisierungstendenzen bei Sprachassistenten können dadurch entstehen, dass die beiden oben beschriebenen Endlosschleifen zu erheblichen Qualitätsunterschieden führen, die es neuen Anbietern von Sprachassistenten erschweren, wettbewerblich gegen die große Nutzerbasis eines etablierten (Öko-)Systems anzukommen. Die Qualitätsunterschiede können ab einem bestimmten Zeitpunkt so erheblich sein, dass die resultierende Qualitätslücke zwischen verschiedenen Sprachassistenten für die Nutzer bzw. Endgerätehersteller deutlich spürbar wird und sich immer mehr Nutzer bzw. Endgerätehersteller für den „besseren“ Sprachassistenten entscheiden. So könnte am Ende ein Monopol entstehen, das in der Lage ist, so stark qualitativ und quantitativ zu skalieren (einschließlich Lernkurven- und Erfahrungseffekte), dass potenzielle Marktneulinge kaum mehr eine realistische Chance haben, in den Markt für digitale Sprachassistenten einzutreten.

Einige Autoren erwarten, dass es zu genau einer solchen Monopolisierung aufgrund der datengetriebenen Größenvorteile sowie Netzwerk- und Feedbackeffekten kommt (Stucke & Ezrachi 2018, Son & Oh 2018, Bentley et al. 2018, Bager et al. 2017). Im Ergebnis würde dies zu einer neuen Gatekeeper-Rolle für ein oder zwei Unternehmen führen, die die dominanten Sprachassistentensysteme kontrollieren.

Diese Gatekeeper-Rolle entsteht insbesondere aufgrund der

- (1) zahlreichen Ausgestaltungs-, Kontroll- und Steuerungsmöglichkeiten seitens der Plattformbetreiber sowie
- (2) effizienten Bündelung von Suche, Shopping, Navigation, sozialen und professionellen Netzwerken, usw.,

die Sprachassistenten (potenziell) auf sich vereinen. Von einem echten „conversational interface“ sind die heutigen Systeme jedoch noch relativ weit entfernt (Porcheron et al. 2018). Trotzdem ist es möglich, dass durch weitere Fortschritte in der Datenverarbei-

tung mit Hilfe neuronaler Netze sowie umfangreichen Kontext- und Nutzungsverlaufdaten, die die Endgeräte an die verschiedenen Dienste übermitteln, ein Paradigmenwechsel von der heute vorherrschenden Input/Output-Logik hin zu einer proaktiven Ansprache durch Sprachassistenten bevorsteht. So ist es vorstellbar, dass durch weitere Personalisierung (Braun et al. 2019) und durch Machine Learning-gestützte Verfahren der Marketinggestaltung (Shaikh 2016) eine neue Art der Beeinflussung von Konsumenten durch Sprachassistenten stattfindet. Zusätzlich würde in einem solchen Szenario die oben schon ausgeführte mögliche Gatekeeper-Rolle von Sprachassistenten nochmals erheblich verstärkt werden. Je mehr sich die primäre Kommunikation zugunsten eines digitalen Assistenten verschiebt, desto weniger wahrscheinlich wird die unabhängige Suche nach Informationen im Internet, die unabhängige Auswahl von Nachrichten, Kunden- und Produkt-Reviews, Preisvergleichen und anderen Möglichkeiten im World Wide Web (WWW).

Diese (Sprach-)Schnittstelle von Mensch und Maschine geht auch mit einer Reduzierung möglicher Wettbewerbsdimensionen für Anbieter von Produkten und Dienstleistungen einher. Die Konkurrenz zwischen Internetdiensten um die Gunst des Nutzers wird bisher in der Regel auf Display-Benutzeroberflächen mit zahlreichen Anzeige-, Ranking- und Listungsoptionen ausgetragen. Bei Sprachassistenten hingegen erfolgt ein erster Vorschlag durch den digitalen Assistenten auf Basis bisher undefinierter, intransparenter oder durch Algorithmen implizit genutzter Kriterien. Die Angabe von möglichen Alternativen beispielsweise eines weiteren Suchergebnisses oder Produkts von anderen Anbietern erfordert eine bewusste Aufforderung durch den Nutzer. Dies gilt zumindest dann, wenn der Sprachassistent nicht mit anderen Benutzeroberflächen kombiniert verwendet wird.

Eine solche Entwicklung kann aus Verbraucherschutzsicht kritisch zu bewerten sein. Ebenso können die in den Abschnitten 2.3.1 und 2.3.2 dargestellten Datenschutz- und Sicherheitsfragen erhöhte Aufmerksamkeit von Seiten des Verbraucherschutzes und der Politik erfordern. In jedem Fall sollten Konsumenten klar und verständlich zu der Verwendung ihrer Daten und möglicher Sicherheitslücken, die auch und gerade durch Fehlbedienung der Systeme entstehen können, informiert werden.

Die Integration zahlreicher wesentlicher Funktionen von bisher separierten Internetdiensten kann Auswirkungen auf Wettbewerb und Verbraucher und damit unterschiedliche Herausforderungen für die Regulierung mit sich bringen:

1. Die zunehmende Bedeutung von Sprachassistenten hat sowohl Auswirkungen auf andere Dienste anderer Anbieter als auch auf die Dienste eines Anbieters innerhalb seines Ökosystems, z. B. Hebelung der Marktmacht, Übertragung auf andere Märkte, Behinderungs-, und Ausbeutungsmisbrauch. Diese Interdependenzen und weiteren ökonomischen Effekte auf dem Markt für Sprachassistenten gilt es in Zukunft zu beobachten, zu untersuchen und einzuordnen.

2. Die (Sprach-)Schnittstelle von Mensch und Maschine kann zukünftig zu einem bedeutenden Machtfaktor im gesamten Internet-Ökosystem werden. Nur Plattformanbieter, die hier erfolgreich sind, sind auch in der Lage, die Bedeutung und Entwicklung ihres digitalen Ökosystems kontrollieren und steuern zu können. Für die Politik wird es hierbei zentral sein, sowohl den wettbewerblichen Rahmen<sup>59</sup> als auch den Daten- und Verbraucherschutz sowie die Datensicherheit zu gewährleisten. Die Notwendigkeit einer reibungslosen Zusammenarbeit der unterschiedlichen Behörden wird damit weiter steigen. Dies schließt insbesondere die Koordinierung und Kooperation auf europäischer und vermehrt globaler Ebene mit ein.
3. Die systembedingte und dauerhafte Verbindung von digitalen Assistenten zur Informationsübermittlung über das Internet macht sie auch anfällig für Manipulationen durch Dritte, z. B. durch Bluetooth-Schwachstellen. Hinzu kommen die hier dargestellten umfangreichen Angriffs- und Eingriffsmöglichkeiten über die Sprachschnittstelle selbst. Dies wirft nicht zuletzt Fragen in Bezug auf die derzeit noch ungeklärte Haftung autonomer Systeme auf. Es gilt sicherzustellen, dass ausschließlich der Besitzer (natürliche bzw. juristische Person) eines Sprachassistenten diesen auch konfigurieren und nutzen kann. Dritte sollen keinen unbefugten Zugang zu Sprachassistenten haben und diese beeinflussen können. Für diese Herausforderung erscheint es zielführend, dass Regulierer und die entsprechenden Anbieter von Sprachassistenten einen passenden (Sicherheits-) Anforderungskatalog einschließlich relevanter Standards erarbeiten und ggf. im Rahmen einer Selbstverpflichtung oder eines Code of Conduct umsetzen.
4. Aufgrund der Verengung des Interaktionsvorgangs von der Display- hin zur Sprachsteuerung gewinnen die zu definierenden Informationspflichten seitens des Sprachassistenten zunehmend an Bedeutung. Diese Informationspflichten gilt es durch den Regulierer für jegliche (auch zukünftige) Art von Diensten und Funktionen zu spezifizieren. Dabei muss einerseits der Zeitaufwand und konkrete Nutzen für den Nutzer verhältnismäßig bleiben (z. B. Bestellprozess im eCommerce) und andererseits sichergestellt werden, dass das Eingrenzen von anzugebenden Alternativen nicht mit negativen Folgewirkungen für Wettbewerb und Verbraucher einhergeht.<sup>60</sup>
5. Multifunktionale Sprachschnittstellen verändern das Nutzungs- und Konsumverhalten sukzessive und damit auch die präferierten Zahlungs- und Abwicklungsstandards (z. B. bargeldloser Zahlungsverkehr) sowie die standardmäßig angebotenen Alternativen. Die Gewährleistung ausreichender Zahlungs- und Abwicklungsalternativen sollte ebenfalls eine wichtige Rolle für die Regulierung spielen.

---

<sup>59</sup> Eine zukünftig zu diskutierende Einflussgröße, die gerade bei Sprachassistenten aufgrund der möglichen Verengung auf die sprachliche Interaktion eine wesentliche Rolle spielen dürfte, ist die sogenannte „Device Neutrality“, mit der sich unter anderem kürzlich veröffentlichte Berichte verschiedener Regulierungsbehörden auseinandersetzen (ACM 2019, Arcep 2018, RTR 2019).

<sup>60</sup> Aus psychologischer Sicht ist hierbei ggf. das sogenannte „Paradox of Choice“ zu beachten (vgl. Schwartz 2004).

## 6 Fazit

Sprachassistenten wie Alexa, Bixby, Cortana, Google Assistant und Siri können in der Zukunft eine bedeutende Rolle im Alltag von Konsumenten in Deutschland und weltweit spielen. Die Ergebnisse des vorliegenden Diskussionsbeitrags unterstreichen das Potenzial der Technologie, die heute schon für etwa 85% der Konsumenten in zumindest einem der Endgeräte, die sie besitzen, verfügbar ist. Unsere Ergebnisse zeigen jedoch genauso die derzeitigen Limitationen von Sprachassistenten sowie Risiken in Bezug auf den Datenschutz und die Sicherheit der Systeme auf. Daraus wurden mögliche Herausforderungen für Wettbewerb und Regulierung abgeleitet.

Trotz der enormen technologischen Fortschritte, die gerade im Bereich des NLP in den letzten Jahren gemacht wurden, und der damit einhergehenden Qualitätsverbesserung von Sprachassistenten, nutzen laut unserer Befragung aktuell nur 26% der Konsumenten in Deutschland heute schon zumindest einen Sprachassistenten. Ein wesentlicher Grund für diesen relativ niedrigen Wert liegt vermutlich in den immer noch mangelnden Konversationsfähigkeiten der Systeme. Ebenso findet sich unter den Befragten ein gewisses Misstrauen gegenüber den hier betrachteten Sprachassistenten. Konsumenten empfinden diese als beeinflussend und verkaufsorientiert. Darüber hinaus haben einige der Befragten Bedenken in Bezug auf die Sammlung und Verwendung ihrer Daten durch ein „ständig lauschendes“ System.

Wie unsere Ergebnisse zeigen, sind diese Sorgen nicht völlig unberechtigt. In der Tat übertragen die populären Systeme nicht nur Informationen, die nach dem sogenannten Wake-Word aufgenommen werden, sondern ebenfalls unterschiedlich viele Informationen, die sich aus den zuvor kontinuierlich gepufferten Daten ergeben. Hier entsteht ein wesentlicher Trade-off für Konsumenten sowie möglicherweise zukünftige regulatorische Rahmenbedingungen. Denn die Verarbeitung von Kontextinformationen ist laut der aktuellen Forschung im Bereich der Mensch-Maschine-Interaktion mithin der entscheidende Faktor für eine hohe Quality of Experience in der Verwendung von Sprachassistenten. Zu restriktive Vorgaben in Bezug auf die Verarbeitung von solchen Informationen können damit die zukünftige Innovation von Sprachassistenten an einem neuralgischen Punkt treffen.

Mit Blick auf den Wettbewerb um die Gunst der Konsumenten zeigt unsere Analyse das zukünftige Risiko eines dominanten Systems oder einer engen Auswahl an dominanten Systemen. Aufgrund der hohen Komplexität der Technologie, die sowohl von umfangreichen und insbesondere passgenauen Nutzerdaten sowie hoher methodischer Kompetenz abhängt, haben Unternehmen mit einem umfangreichen Ökosystem an Endgeräten und Diensten einen deutlichen komparativen Vorteil gegenüber neuen Mitbewerbern.

Gerade mit Blick auf die große vorinstallierte Basis an Endgeräten, die ohne weitere Ausgaben oder Anstrengung der Nutzer Sprachassistentensysteme bereitstellen, sollten Regulierer und die Politik die Entwicklung der Systeme sowie die Nutzungsmuster der Konsumenten kontinuierlich beobachten.

## Literaturverzeichnis

- ACM. 2019. Market study into mobile app stores. Den Haag: The Netherlands Authority for Consumers & Markets.
- Alepis, Efthimos, & Constantinos Patsakis. 2017. "Monkey Says, Monkey Does: Security and Privacy on Voice Assistants." *IEEE Access* 5:17841-17851.
- Ammari, Tawfiq, Jofish Kaye, Janice Y. Tsai, & Frank Bentley. 2019. "Music, Search, and IoT: How People (Really) Use Voice Assistants." *ACM Transactions on Computer-Human Interaction (TOCHI)* 26 (3):17:1-17:29.
- Arcep. 2018. Smartphones, tablets, voice assistants...Devices, the weak link in achieving an open internet - Report on their limitations and proposals for corrective measures.
- Armstrong, Mark. 2006. "Competition in Two-Sided Markets." *The RAND Journal of Economics* 37 (3):668-691.
- Arnold, René, Martin Waldburger, & Annette Hillebrand. 2015. Personal Data and Privacy - A Study for Ofcom. Bad Honnef: WIK-Consult.
- Azvine, Behnam, David Djan, Kwok Ching Tsui, & Wayne Wobcke. 2000. "The Intelligent Assistant: An Overview." In *Intelligent Systems and Soft Computing - Prospects, Tools and Applications*, edited by Benham Azvine, Detlef D. Nauck & Nader Azarmi, 215-238. Berlin, Heidelberg: Springer.
- Baber, Christopher. 1993. "Developing Interactive Speech Technology." In *Interactive Speech technology - Human Factors Issues in the Application of Speech Input/Output to Computers*, edited by Christopher Baber & Janet M. Noyes, 1-18. New York: Taylor & Francis.
- Bager, Jo, Nico Jurrán, Holger Bleich, Martin Reche, Jörg Heidrich, & Nicolas Maekeler. 2017. "Assistent allgegenwärtig. Digitale Assistenten vom Spielzeug für Nerds zur Bedienoberfläche für alles." *c't - Magazin für Computer und Technik*, 64-86.
- Barth, Susanne, & Menno D. T. de Jong. 2017. "The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review." *Telematics and Informatics* 34 (7):1038-1058.
- Benbasat, Izak, & Weiquan Wang. 2005. "Trust In and Adoption of Online Recommendation Agents." *Journal of the Association for Information Systems* 6 (3):72-101.
- Beneteau, Erin, Olivia K. Richards, Mingrui Zhang, Julie A. Kientz, Jason Yip, & Alexis Hiniker. 2019. "Communication Breakdowns Between Families and Alexa." 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, 04.-09. May 2019.
- Bentley, Frank, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, & Danielle Lottridge. 2018. "Understanding the Long-Term Use of Smart Speaker Assistants." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2 (3):91-105.
- Bisher Tarakji, Ahmad Bisher, Jian Xu, Juan A. Colmenares, & Iqbal Mohamed. 2018. "Voice Enabling Mobile Applications with UIVoice." 1st International Workshop on Edge Systems, Analytics and Networking, Munich, 10.-15. June 2018.
- Bispham, Mary K., Ioannis Agrafiotis, & Michael Goldsmith. 2018. "A Taxonomy of Attacks via the Speech Interface." Conference on Cyber-Technologies and Cyber-Systems, Athens, 18.-22. November 2018.
- Bispham, Mary K., Ioannis Agrafiotis, & Michael Goldsmith. 2019. "Nonsense Attacks on Google Assistant and Missense Attacks on Amazon Alexa." 5th International Conference on Information Systems, Security and Privacy, Prague, 23.-25. February 2019.

- Bohm, Allison S., Edward J. George, Bennett Cyphers, & Shirley Lu. 2017. "Privacy and Liberty in an Always-on, Always-Listening World." *The Columbia Science & Technology Law Review* 19:1-45.
- Braun, Michael, Anja Mainz, Ronee Chadowitz, Bastian Pfleging, & Florian Alt. 2019. "At your service: Designing voice assistant personalities to improve automotive user interfaces." 2019 CHI Conference on Human Factors in Computing Systems, Glasgow.
- Budzinski, Oliver, Victoriia Noskova, & Xijie Zhang. 2018. "The Brave New World of Digital Personal Assistants: Benefits and Challenges from an Economic Perspective." *Ilmenau Economics Discussion Papers* 24 (118):1-30.
- Buvat, Jerome, Mark Taylor, Kees Jacobs, Amol Khadikar, & Amrita Sengupta. 2018. *Conversational Commerce: Why Consumers Are Embracing Voice Assistants in Their Lives*. Capgemini Research Institute.
- Carlini, Nicholas, Pratyush Mishra, Tavish Vaidya, Yuankai Zhang, Micah Sherr, Clay Shields, David Wagner, & Wenchao Zhou. 2016. "Hidden Voice Commands." 25th USENIX Conference on Security Symposium, Austin, TX, 10.-12. August 2016.
- Chen, Hongshen, Xiaorui Liu, Dawei Yin, & Jiliang Tang. 2017. "A Survey on Dialogue Systems: Recent Advances and New Frontiers." *ACM SIGKDD Explorations Newsletter* 19 (2):25-35.
- Chen, Mei-Ling, & Hao-Chuan Wang. 2018. "How Personal Experience and Technical Knowledge Affect Using Conversational Agents." 23rd International Conference on Intelligent User Interfaces Companion, Tokyo, 07.-11. March 2018.
- Chung, Hyunji, Michaela Iorga, Jeffrey Voas, & Sangjin Lee. 2017. "Alexa, Can I Trust You?" *Computer* 50 (9):100-104.
- Chung, Hyunji, Jungheum Park, & Sangjin Lee. 2017. "Digital forensic approaches for Amazon Alexa ecosystem." *Digital Investigation* 22:S15-S25.
- Clark, Leigh, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, & Benjamin R. Cowan. 2019. "What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents." Conference on Human Factors in Computing Systems, Glasgow, 04.-09. May 2019.
- Cohen, Michael H., James P. Giangola, & Jennifer Balogh. 2004. *Voice user interface design*. Boston: Addison-Wesley Professional.
- Cowan, Benjamin R., Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, & Natasha Bandeira. 2017. "'What Can I Help You With?': Infrequent User's Experiences of Intelligent Personal Assistants." 19th International Conference on Human-Computer Interaction with Mobile Devices and Services Vienna, 04.-07. September 2017.
- Diao, Wenrui, Xiangyu Liu, Zhe Zhou, & Kehuan Zhang. 2014. "Your Voice Assistant is Mine: How to Abuse Speakers to Steal Information and Control Your Phone." 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices, Scottsdale, 07. November 2014.
- Dubiel, Mateusz, Martin Halvey, & Leif Azzopardi. 2019. "A Survey Investigating Usage of Virtual Personal Assistants." *Computing Research Repository*.
- Fraser, Norman. 1998. "Assessment of interactive systems." In *Handbook of Standards and Resources for Spoken Language Systems*, edited by Dafydd Gibbon, Roger Moore & Richard Winski, 564-615. Berlin: Mouton de Gruyter.
- Han, Sangyeal, & Heetae Yang. 2018. "Understanding Adoption of Intelligent Personal Assistants: A Parasocial Relationship Perspective." *Industrial Management & Data Systems* 118 (3):618-636.



- Hashemi, Seyyed H., Kyle Williams, Ahmed E El Kholy, Imed Zitouni, & Paul A. Cook. 2018. "Measuring User Satisfaction on Smart Speaker Intelligent Assistants Using Intent Sensitive Query Embeddings." International Conference on Information and Knowledge Management Torino, 22.-26. October 2018.
- Hauswald, Johann, Michael A. Laurenzano, Yunqi Zhang, Cheng Li, Austin Rovinski, Arjun Khurana, Ronald G. Dreslinski, Trevor Mudge, Vinicius Petrucci, Lingjia Tang, & Jason Mars. 2015. "Sirius: An Open End-to-End Voice and Vision Personal Assistant and Its Implications for Future Warehouse Scale Computers." International Conference on Architectural Support for Programming Languages and Operating Systems, Istanbul, 14.-18. March 2015.
- Hoy, Matthew B. 2018. "Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants." *Medical Reference Services Quarterly* 37 (1):81-88.
- Huang, Xuedong, James Baker, & Raj Reddy. 2014. "A Historical Perspective of Speech Recognition." *Communications of the ACM* 57 (1):94-103.
- Jiang, Jiepu, Wei Jeng, & Daqing He. 2013. "How Do Users Respond to Voice Input Errors?: Lexical and Phonetic Query Reformulation in Voice Search." 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, Dublin, 28. July - 01. August 2013.
- Jokinen, Kristiina, & Michael McTear. 2009. "Spoken Dialogue Systems " *Synthesis Lectures on Human Language Technologies* 2 (1):1-151.
- Kinsella, Bret, & Ava Mutchler. 2018. Voice Assistant Consumer Adoption Report. voicebot.ai, PullString, RAIN.
- Kiseleva, Julia, Kyle Williams, Jiepu Jiang, Ahmed H. Awadallah, Aidan C. Crook, Imed Zitouni, & Tasos Anastasakos. 2016. "Understanding User Satisfaction with Intelligent Assistants." Conference on Human Information Interaction and Retrieval, Carrboro, 13.-17. March 2016.
- Knote, Robin, Andreas Janson, Laura Eigenbrod, & Matthias Söllner. 2018. "The What and How of Smart Personal Assistants: Principles and Applications for IS Research." Multikonferenz Wirtschaftsinformatik (MKWI), Lüneburg, 06. March 2018.
- Knote, Robin, Andreas Janson, Matthias Schnöller, & Jan M. Leimeister. 2019. "Classifying Smart Personal Assistants: An Empirical Cluster Analysis." 52nd Hawaii International Conference on System Sciences, Hawaii, 08.-11. January 2019.
- Kopp, Stefan, Lars Gesellensetter, Nicole C. Krämer, & Ipke Wachsmuth. 2005. "A Conversational Agent as Museum Guide - Design and Evaluation of a Real-World Application." In *International Workshop on Intelligent Virtual Agents*, 329-343. Berlin, Heidelberg: Springer.
- Kumar, Deepak, Riccardo Paccagnella, Paul Murley, Eric Hennenfent, Joshua Mason, Adam Bates, & Michael Bailey. 2018. "Skill Squatting Attacks on Amazon Alexa." 27th USENIX Security Symposium, Baltimore, 15.-17. August 2018.
- Lau, Josephine, Benjamin Zimmerman, & Florian Schaub. 2018. "Alexa, Are You Listening?: Privacy Perceptions, Concerns and Privacy-seeking Behaviors with Smart Speakers." *Proceedings of the ACM on Human-Computer Interaction - CSCW 2 (CSCW)*:Article No. 102.
- Leong, Lee H., Shinsuke Kobayashi, Noboru Koshizuka, & Ken Sakamura. 2005. "CASIS: A Context-Aware Speech Interface System." 10th International Conference on Intelligent User Interfaces, San Diego, 10.-13. January 2005.

- Liao, Yuting, Jessica Vitak, Priya Kumar, Michael Zimmer, & Katherine Kritikos. 2019. "Understanding the Role of Privacy and Trust in Intelligent Personal Assistant Adoption." In *Information in Contemporary Society*, edited by Natalie Greene Taylor, Caitlin Christian-Lamb, Michelle H. Martin & Bonnie Nardi, 102-113. Berlin, Heidelberg: Springer.
- Lopatovska, Irene, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorche, David Hirsch, Qi Li, & Adrianna Martinez. 2018. "Talk to Me: Exploring User Interaction with the Amazon Alexa." *Journal of Librarianship and Information Science* 00 (0):1-14.
- López-Cózar, Ramón, Ángel de la Torre, José C. Segura, & Antonio J. Rubio. 2003. "Assessment of Dialogue Systems by Means of a New Simulation Technique." *Speech Communication* 40 (3):387-407.
- Luger, Ewa, & Abigail Sellen. 2016. "'Like Having a Really Bad PA': The Gulf between User Expectation and Experience of Conversational Agents." 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, 07.-12. May 2016.
- Lutter, Timm, Angelika Pentsi, Michael Poguntke, Klaus Böhm, & Ralf Esser. 2015. *Zukunft der Consumer Electronics – 2015: Marktentwicklung, Schlüsseltrends, Konsumentenverhalten, Mediennutzung, Neue Technologien*. Berlin: Deloitte, Bitkom.
- McTear, Michael. 2002. "Spoken Dialogue Technology: Enabling the Conversational User Interface." *ACM Computing Surveys* 34 (1):90-169.
- McTear, Michael, Zoraida Callejas, & David Griol. 2016. *The Conversational Interface: Talking to Smart Devices*. Cham: Springer International Publishing.
- Myers, Chelsea, Anushay Furqan, Jessica Nebolsky, Karina Caro, & Jichen Zhu. 2018. "Patterns for How Users Overcome Obstacles in Voice User Interfaces." 2018 CHI Conference on Human Factors in Computing Systems, Montreal, 21.-26. April 2018.
- Nass, Clifford, & Youngme Moon. 2000. "Machines and Mindlessness: Social Responses to Computers." *Journal of Social Issues* 56 (1):81-103.
- Newman, Nic. 2018. *The Future of Voice and the Implications for News*. Reuters Insitute, University of Oxford.
- Olaso Fernández, Javier M. 2017. "Spoken Dialogue Systems: Architectures and Applications." PhD, Universidad del Pais Vasco.
- Pelikan, Hannah R. M., & Mathias Broth. 2016. "Why That Nao?: How Humans Adapt to a Conventional Humanoid Robot in Taking Turns-at-Talk." 2016 CHI Conference on Human Factors in Computing Systems, San Jose, 07.-12. May 2016.
- Porcheron, Martin, Joel E. Fischer, Stuart Reeves, & Sarah Sharples. 2018. "Voice Interfaces in Everyday Life." 2018 CHI Conference on Human Factors in Computing Systems, Montreal, 21.-26. April 2018.
- Purington, Amanda, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, & Samuel Hardman Taylor. 2017. "'Alexa is my new BFF': Social Roles, User Satisfaction, and Personification of the Amazon Echo." 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, 06.-11. May 2017.
- PwC. 2017. *Digitale Assistenten - Bevölkerungsbefragung 2017*. Düsseldorf.
- PwC. 2018. *Consumer Intelligence Series: Prepare for the Voice Revolution*.
- Reeves, Stuart, Martin Porcheron, & Joel Fischer. 2018. "'This Is Not What We Wanted': Designing for Conversation with Voice Interfaces." *Interactions* 26 (1):46-51.
- Rogers, Everett M. 1962. *Diffusion of Innovations*. New York: The Free Press.

- Rong, Xin, Adam Fourney, Robin N. Brewer, Meredith Ringel Morris, & Paul N. Bennett. 2017. "Managing Uncertainty in Time Expressions for Virtual Assistants." 2017 CHI Conference on Human Factors in Computing Systems, Denver, 06.-11. May 2017.
- RTR. 2019. Zur Offenheit des Internets - Betriebssysteme, Apps und App Stores. Wien: Rundfunk und Telekom Regulierungs-GmbH.
- Sarikaya, Ruhi. 2017. "The Technology Behind Personal Digital Assistants: An Overview of the System Architecture and Key Components." *IEEE Signal Processing Magazine* 34 (1):67-81.
- Schwartz, Barry. 2004. "The paradox of choice: Why more is less."
- Sciuto, Alex, Arnita Saini, Jodi Forlizzi, & Jason I. Hong. 2018. "'Hey Alexa, What's Up?': A Mixed-Methods Studies of In-Home Conversational Agent Usage." 2018 Designing Interactive Systems Conference, Hong Kong, 09.-13. June 2018.
- Serban, Iulian V., Ryan Lowe, Peter Henderson, Laurent Charlin, & Joelle Pineau. 2018. "A Survey of Available Corpora for Building Data-Driven Dialogue Systems: The Journal Version." *Dialogue & Discourse* 9 (1):1-49.
- Shaikh, Samira. 2016. "Persuasion in Online Communication - Automation and Counteraction." PhD, Department of Computer Science, State University of New York.
- Son, Yoonseock, & Wonseok Oh. 2018. "Alexa, Buy Me a Movie!: How AI Speakers Reshape Digital Content Consumption and Preference." International Conference on Information Systems 2018, San Fransisco, 13.-16. December 2018.
- Song, Liwei, & Prateek Mittal. 2017. "Inaudible Voice Commands." 2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, 30. October - 03. November 2017.
- Statistisches Bundesamt (Destatis). 2019. Private Haushalte in der Informationsgesellschaft - Nutzung von Informations- und Kommunikationstechnologien. In *Wirtschaftsrechnungen*. Bonn.
- Stucke, Maurice E., & Ariel Ezrachi. 2018. "Alexa et al., What Are You Doing with My Data?" *Critical Analysis of Law* 5 (1):148-169.
- Taylor, Mark, Kees Jacobs, Shannon Warner, Gwendolyn Graman, Sanjay Sarma, Brian Subirana, Richard Cantwell, Chris Hunt, & Jon Stine. 2017. Time to Talk: The Future for Brands in Conversational. Capgemini, intel, MIT, Auto-ID Lab.
- von der Pütten, Astrid M., Nicole C. Krämer, Jonathan Gratch, & Sin-Hwa Kang. 2010. "'It doesn't matter what you are!' Explaining social effects of agents and avatars." *Computers in Human Behavior* 26 (6):1641-1650.
- Wagner, Gunther, Andreas Gentner, Michael Müller, Thomas Schlaak, Ralf Esser, Thomas Nugel, & Dominik Busching. 2018. Smart Home Consumer Survey 2018: Ausgewählte Ergebnisse für den deutschen Markt. Deloitte.
- Wang, Xin, Shinji Takaki, & Junichi Yamagishi. 2016. "A Comparative Study of the Performance of HMM, DNN, and RNN based Speech Synthesis Systems Trained on Very Large Speaker-Dependent Corpora." 9th ISCA Speech Synthesis Workshop, Sunnydale, 13.-15. September 2016.
- Weinschenk, Susan, & Dean T. Barker. 2000. *Designing Effective Speech Interfaces*. New York: Wiley.
- Wyard, Peter J., Andre D. Simons, Steve Appleby, Ed Kaneen, Sandra H. Williams, & Keith R. Preston. 1996. "Spoken Language Systems: Beyond Prompt and Response." *BT Technology Journal* 14 (1):187-205.

- Xu, Shanchuan, Wanlin Gao, Zhen Li, Shuliang Zhang, & Jianing Zhao. 2010. "Design of Hierarchical and Configurable IVR System." 2010 Second International Conference on Computational Intelligence and Natural Computing, Wuhan, 13.-14. September 2010.
- Yarosh, Svetlana, Stryker Thompson, Kathleen Watson, Alice Chase, Ashwin Senthilkumar, Ye Yuan, & A. J. Bernheim Brush. 2018. "Children Asking Questions: Speech Interface Reformulations and Personification Preferences." 17th ACM Conference on Interaction Design and Children, Trondheim, 19.-22. June 2018.
- Zhang, Guoming, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, & Wenyuan Xu. 2017. "DolphinAttack: Inaudible Voice Commands." 2017 ACM SIGSAC Conference on Computer and Communications Security, Dallas, 30. October - 03. November 2017.
- Zhang, Yangyong, Lei Xu, Abner Mendoza, Guangliang Yang, Phakpoom Chinprutthiwong, & Guofe Gu. 2019. "Life after Speech Recognition: Fuzzing Semantic Misinterpretation for Voice Assistant Applications." Network and Distributed Systems Security (NDSS) Symposium, San Diego, 24.-27. February 2019.
- Zue, Victor W., & James R. Glass. 2000. "Conversational Interfaces: Advances and Challenges." *Proceedings of the IEEE* 88 (8):1166-1180.

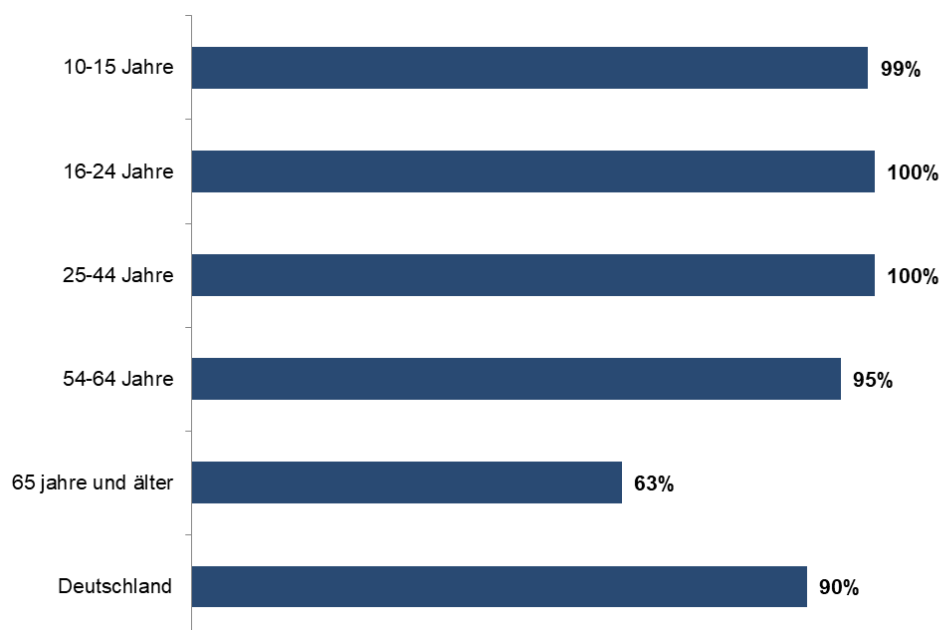
## Anhang: Methodik

Die in dieser Arbeit verwendeten Ergebnisse basieren wesentlich auf einer im November 2018 durchgeführten Online-Verbraucherbefragung des WIK. Anhand dieser wurde die Entwicklung der Nutzung von Sprachassistenten untersucht. Die Befragung wurde mit Computer Aided Web Interviewing (CAWI) erfasst. Die Stichprobengröße lag bei 3184 Befragten.

Es handelt sich um keine reine Zufallsstichproben, da Quoten für Alter und Geschlecht basierend auf der Verteilung der Bevölkerung in Deutschland festgelegt wurden. Daher handelt es sich hier speziell um ein Quotenstichprobe, um eine bevölkerungsrepräsentative Zusammenstellung zu gewährleisten, in diesem Fall für die deutsche Bevölkerung ab 18 Jahren.

Grundsätzlich decken Online-Befragungen nur den Teil der Bevölkerung ab, der das Internet nutzt bzw. Zugang zu diesem hat – also somit die Internetnutzer. In Deutschland sind dies laut des Statistischen Bundesamts etwa 90% der Bevölkerung. Die meisten, die kein Internet nutzen, stammen aus der Altersgruppe 65+. Die Anteile der Nutzer nach Altersgruppen sind in Abbildung A-1 zusammengefasst. Hier wird ersichtlich, dass nahezu alle Personen bis 44 Jahre in Deutschland das Internet benutzen. In der Altersgruppe 45 bis 64 Jahre sind es immerhin noch 95%. Lediglich Personen älter als 65 Jahre nutzen das Internet weniger: Nur 63% von diesen nutzen es.

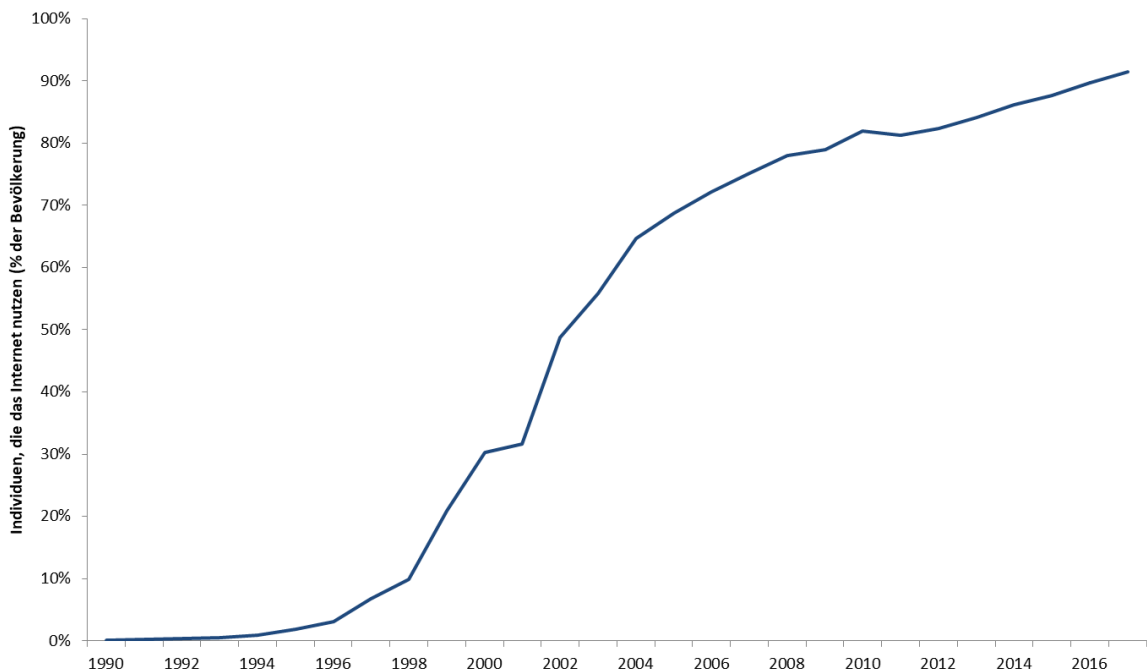
Abbildung A-1: Internetnutzung nach Altersgruppen (2018)



Quelle: Eigene Darstellung basierend auf Daten von Statistisches Bundesamt (Destatis) (2019).

In unserer Online-Befragung können wir daher mit Hilfe der Quotenstichprobe das Nutzungsverhalten im Hinblick auf Sprachassistenten zumindest für die Altersgruppe bis 54 Jahre bevölkerungsrepräsentativ abbilden. Bei den Ergebnissen der Altersgruppe 55+ besteht die Möglichkeit, dass eine gewisse Überschätzung in den Nutzungsmustern und den Nutzeranteilen von Sprachassistenten vorliegt.

Abbildung A-2: Entwicklung der Internetnutzung in Deutschland – Bevölkerungsanteil



Quelle: Eigene Darstellung; Daten stammen von der Weltbank und dem Statistischen Bundesamt.



Als "Diskussionsbeiträge" des Wissenschaftlichen Instituts für Infrastruktur und Kommunikationsdienste sind zuletzt erschienen:

- Nr. 361: Alex Kalevi Dieke, Petra Junk, Martin Zauner:  
Qualitätsfaktoren in der Post-Entgeltregulierung, November 2011
- Nr. 362: Gernot Müller:  
Die Bedeutung von Liberalisierungs- und Regulierungsstrategien für die Entwicklung des Eisenbahnpersonenfernverkehrs in Deutschland, Großbritannien und Schweden, Dezember 2011
- Nr. 363: Wolfgang Kiesewetter:  
Die Empfehlungspraxis der EU-Kommission im Lichte einer zunehmenden Differenzierung nationaler Besonderheiten in den Wettbewerbsbedingungen unter besonderer Berücksichtigung der Relevante-Märkte-Empfehlung, Dezember 2011
- Nr. 364: Christine Müller, Andrea Schweinsberg:  
Vom Smart Grid zum Smart Market – Chancen einer plattformbasierten Interaktion, Januar 2012
- Nr. 365: Franz Büllingen, Annette Hillebrand, Peter Stamm, Anne Stetter:  
Analyse der Kabelbranche und ihrer Migrationsstrategien auf dem Weg in die NGA-Welt, Februar 2012
- Nr. 366: Dieter Elixmann, Christin-Isabel Gries, J. Scott Marcus:  
Netzneutralität im Mobilfunk, März 2012
- Nr. 367: Nicole Angenendt, Christine Müller, Marcus Stronzik:  
Elektromobilität in Europa: Ökonomische, rechtliche und regulatorische Behandlung von zu errichtender Infrastruktur im internationalen Vergleich, Juni 2012
- Nr. 368: Alex Kalevi Dieke, Petra Junk, Sonja Thiele, Martin Zauner:  
Kostenstandards in der Ex-Post-Preiskontrolle im Postmarkt, Juni 2012
- Nr. 369: Ulrich Stumpf, Stefano Lucidi:  
Regulatorische Ansätze zur Vermeidung wettbewerbswidriger Wirkungen von Triple-Play-Produkten, Juni 2012
- Nr. 370: Matthias Wissner:  
Marktmacht auf dem Primär- und Sekundär-Regelenergiemarkt, Juli 2012
- Nr. 371: Antonia Niederprüm, Sonja Thiele:  
Prognosemodelle zur Nachfrage von Briefdienstleistungen, Dezember 2012
- Nr. 372: Thomas Plückebaum, Matthias Wissner:  
Bandbreitenbedarf für Intelligente Stromnetze, 2013
- Nr. 373: Christine Müller, Andrea Schweinsberg:  
Der Netzbetreiber an der Schnittstelle von Markt und Regulierung, 2013
- Nr. 374: Thomas Plückebaum:  
VDSL Vectoring, Bonding und Phantoming: Technisches Konzept, marktliche und regulatorische Implikationen, Januar 2013
- Nr. 375: Gernot Müller, Martin Zauner:  
Einzelwagenverkehr als Kernelement eisenbahnbezogener Güterverkehrskonzepte?, Dezember 2012
- Nr. 376: Christin-Isabel Gries, Imme Philbeck:  
Marktentwicklungen im Bereich Content Delivery Networks, April 2013
- Nr. 377: Alessandro Monti, Ralf Schäfer, Stefano Lucidi, Ulrich Stumpf:  
Kundenbindungsansätze im deutschen TK-Markt im Lichte der Regulierung, Februar 2013
- Nr. 378: Tseveen Gantumur:  
Empirische Erkenntnisse zur Breitbandförderung in Deutschland, Juni 2013
- Nr. 379: Marcus Stronzik:  
Investitions- und Innovationsanreize: Ein Vergleich zwischen Revenue Cap und Yardstick Competition, September 2013
- Nr. 380: Dragan Ilic, Stephan Jay, Thomas Plückebaum, Peter Stamm:  
Migrationsoptionen für Breitbandkabelnetze und ihr Investitionsbedarf, August 2013



- Nr. 381: Matthias Wissner:  
Regulierungsbedürftigkeit des Fernwärmesektors, Oktober 2013
- Nr. 382: Christian M. Bender, Alex Kalevi Dieke, Petra Junk, Sonja Thiele:  
Netzzugang im Briefmarkt, Oktober 2013
- Nr. 383: Andrea Liebe, Christine Müller:  
Energiegenossenschaften im Zeichen der Energiewende, Januar 2014
- Nr. 384: Christian M. Bender, Marcus Stronzik:  
Verfahren zur Ermittlung des sektoralen Produktivitätsfortschritts - Internationale Erfahrungen und Implikationen für den deutschen Eisenbahninfrastruktursektor, März 2014
- Nr. 385: Franz Büllingen, Annette Hillebrand, Peter Stamm:  
Die Marktentwicklung für Cloud-Dienste - mögliche Anforderungen an die Netzinfrastruktur, April 2014
- Nr. 386: Marcus Stronzik, Matthias Wissner:  
Smart Metering Gas, März 2014
- Nr. 387: René Arnold, Sebastian Tenbrock:  
Bestimmungsgründe der FTTP-Nachfrage, August 2014
- Nr. 388: Lorenz Nett, Stephan Jay:  
Entwicklung dynamischer Marktszenarien und Wettbewerbskonstellationen zwischen Glasfasernetzen, Kupfernetzen und Kabelnetzen in Deutschland, September 2014
- Nr. 389: Stephan Schmitt:  
Energieeffizienz und Netzregulierung, November 2014
- Nr. 390: Stephan Jay, Thomas Plückerbaum:  
Kostensenkungspotenziale für Glasfaseranschlussnetze durch Mitverlegung mit Stromnetzen, September 2014
- Nr. 391: Peter Stamm, Franz Büllingen:  
Stellenwert und Marktperspektiven öffentlicher sowie privater Funknetze im Kontext steigender Nachfrage nach nomadischer und mobiler hochbitratiger Datenübertragung, Oktober 2014
- Nr. 392: Dieter Elixmann, J. Scott Marcus, Thomas Plückerbaum:  
IP-Netzzusammenschaltung bei NGN-basierten Sprachdiensten und die Migration zu All-IP: Ein internationaler Vergleich, November 2014
- Nr. 393: Stefano Lucidi, Ulrich Stumpf:  
Implikationen der Internationalisierung von Telekommunikationsnetzen und Diensten für die Nummernverwaltung, Dezember 2014
- Nr. 394: Rolf Schwab:  
Stand und Perspektiven von LTE in Deutschland, Dezember 2014
- Nr. 395: Christian M. Bender, Alex Kalevi Dieke, Petra Junk, Antonia Niederprüm:  
Produktive Effizienz von Postdienstleistern, November 2014
- Nr. 396: Petra Junk, Sonja Thiele:  
Methoden für Verbraucherbefragungen zur Ermittlung des Bedarfs nach Post-Universaldienst, Dezember 2014
- Nr. 397: Stephan Schmitt, Matthias Wissner:  
Analyse des Preissetzungsverhaltens der Netzbetreiber im Zähl- und Messwesen, März 2015
- Nr. 398: Annette Hillebrand, Martin Zauner:  
Qualitätsindikatoren im Brief- und Paketmarkt, Mai 2015
- Nr. 399: Stephan Schmitt, Marcus Stronzik:  
Die Rolle des generellen X-Faktors in verschiedenen Regulierungsregimen, Juli 2015
- Nr. 400: Franz Büllingen, Solveig Börnsen:  
Marktorganisation und Marktrealität von Machine-to-Machine-Kommunikation mit Blick auf Industrie 4.0 und die Vergabe von IPv6-Nummern, August 2015
- Nr. 401: Lorenz Nett, Stefano Lucidi, Ulrich Stumpf:  
Ein Benchmark neuer Ansätze für eine innovative Ausgestaltung von Frequenzgebühren und Implikationen für Deutschland, November 2015

- Nr. 402: Christian M. Bender, Alex Kalevi Dieke, Petra Junk:  
Zur Marktabgrenzung bei Kurier-, Paket- und Expressdiensten, November 2015
- Nr. 403: J. Scott Marcus, Christin Gries, Christian Wernick, Imme Philbeck:  
Entwicklungen im internationalen Mobile Roaming unter besonderer Berücksichtigung struktureller Lösungen, Januar 2016
- Nr. 404: Karl-Heinz Neumann, Stephan Schmitt, Rolf Schwab unter Mitarbeit von Marcus Stronzik:  
Die Bedeutung von TAL-Preisen für den Aufbau von NGA, März 2016
- Nr. 405: Caroline Held, Gabriele Kulenkampff, Thomas Plückebaum:  
Entgelte für den Netzzugang zu staatlich geförderter Breitband-Infrastruktur, März 2016
- Nr. 406: Stephan Schmitt, Matthias Wissner:  
Kapazitätsmechanismen – Internationale Erfahrungen, April 2016
- Nr. 407: Annette Hillebrand, Petra Junk:  
Paketshops im Wettbewerb, April 2016
- Nr. 408: Tseveen Gantumur, Iris Henseler-Unger, Karl-Heinz Neumann:  
Wohlfahrtsökonomische Effekte einer Pure LRIC - Regulierung von Terminierungsentgelten, Mai 2016
- Nr. 409: René Arnold, Christian Hildebrandt, Martin Waldburger:  
Der Markt für Over-The-Top Dienste in Deutschland, Juni 2016
- Nr. 410: Christian Hildebrandt, Lorenz Nett:  
Die Marktanalyse im Kontext von mehrseitigen Online-Plattformen, Juni 2016
- Nr. 411: Tseveen Gantumur, Ulrich Stumpf:  
NGA-Infrastrukturen, Märkte und Regulierungsregime in ausgewählten Ländern, Juni 2016
- Nr. 412: Alex Dieke, Antonia Niederprüm, Sonja Thiele:  
UPU-Endvergütungen und internationaler E-Commerce, September 2016 (in deutscher und englischer Sprache verfügbar)
- Nr. 413: Sebastian Tenbrock, René Arnold:  
Die Bedeutung von Telekommunikation in intelligent vernetzten PKW, Oktober 2016
- Nr. 414: Christian Hildebrandt, René Arnold:  
Big Data und OTT-Geschäftsmodelle sowie daraus resultierende Wettbewerbsprobleme und Herausforderungen bei Datenschutz und Verbraucherschutz, November 2016
- Nr. 415: J. Scott Marcus, Christian Wernick:  
Ansätze zur Messung der Performance im Best-Effort-Internet, November 2016
- Nr. 416: Lorenz Nett, Christian Hildebrandt:  
Marktabgrenzung und Marktmacht bei OTT-0 und OTT-1-Diensten, Eine Projektskizze am Beispiel von Instant-Messenger-Diensten, Januar 2017
- Nr. 417: Peter Kroon:  
Maßnahmen zur Verhinderung von Preis-Kosten-Scheren für NGA-basierte Dienste, Juni 2017
- Nr. 419: Stefano Lucidi:  
Analyse marktstruktureller Kriterien und Diskussion regulatorischer Handlungsoptionen bei engen Oligopolen, April 2017
- Nr. 420: J. Scott Marcus, Christian Wernick, Tseveen Gantumur, Christin Gries:  
Ökonomische Chancen und Risiken einer weitreichenden Harmonisierung und Zentralisierung der TK-Regulierung in Europa, Juni 2017
- Nr. 421: Lorenz Nett:  
Incentive Auctions als ein neues Instrument des Frequenzmanagements, Juli 2017
- Nr. 422: Christin Gries, Christian Wernick:  
Bedeutung der embedded SIM (eSIM) für Wettbewerb und Verbraucher im Mobilfunkmarkt, August 2017

- Nr. 423: Fabian Queder, Nicole Angenendt, Christian Wernick:  
Bedeutung und Entwicklungsperspektiven von öffentlichen WLAN-Netzen in Deutschland, Dezember 2017
- Nr. 424: Stefano Lucidi, Bernd Sörries, Sonja Thiele:  
Wirksamkeit sektorspezifischer Verbraucherschutzregelungen in Deutschland, Januar 2018
- Nr. 425: Bernd Sörries, Lorenz Nett:  
Frequenzpolitische Herausforderungen durch das Internet der Dinge - künftiger Frequenzbedarf durch M2M-Kommunikation und frequenzpolitische Handlungsempfehlungen, März 2018
- Nr. 426: Saskja Schäfer, Gabriele Kulenkampff, Thomas Plückebaum unter Mitarbeit von Stephan Schmitt:  
Zugang zu gebäudeinterner Infrastruktur und adäquate Bepreisung, April 2018
- Nr. 427: Christian Hildebrandt, René Arnold:  
Marktbeobachtung in der digitalen Wirtschaft – Ein Modell zur Analyse von Online-Plattformen, Mai 2018
- Nr. 428: Christin Gries, Christian Wernick:  
Treiber und Hemmnisse für kommerziell verhandelten Zugang zu alternativen FTTB/H-Netzinfrastrukturen, Juli 2018
- Nr. 429: Serpil Taş, René Arnold:  
Breitbandinfrastrukturen und die künftige Nutzung von audiovisuellen Inhalten in Deutschland: Herausforderungen für Kapazitätsmanagement und Netzneutralität, August 2018
- Nr. 430: Sebastian Tenbrock, Sonia Strube Martins, Christian Wernick, Fabian Queder, Iris Henseler-Unger:  
Co-Invest Modelle zum Aufbau von neuen FTTB/H-Netzinfrastrukturen, August 2018
- Nr. 431: Johanna Bott, Christian Hildebrandt, René Arnold:  
Die Nutzung von Daten durch OTT-Dienste zur Abschöpfung von Aufmerksamkeit und Zahlungsbereitschaft: Implikationen für Daten- und Verbraucherschutz, Oktober 2018
- Nr. 432: Petra Junk, Antonia Niederprüm:  
Warenversand im Briefnetz, Oktober 2018
- Nr. 433: Christian M. Bender, Annette Hillebrand:  
Auswirkungen der Digitalisierung auf die Zustellogistik, Oktober 2018
- Nr. 434: Antonia Niederprüm:  
Hybridpost in Deutschland, Oktober 2018
- Nr. 436: Petra Junk:  
Digitalisierung und Briefsubstitution: Erfahrungen in Europa und Schlussfolgerungen für Deutschland, Oktober 2018
- Nr. 437: Peter Kroon, René Arnold:  
Die Bedeutung von Interoperabilität in der digitalen Welt – Neue Herausforderungen in der interpersonellen Kommunikation, Dezember 2018
- Nr. 438: Stefano Lucidi, Bernd Sörries:  
Auswirkung von Bündelprodukten auf den Wettbewerb, März 2019
- Nr. 439: Christian M. Bender, Sonja Thiele:  
Der deutsche Postmarkt als Infrastruktur für europäischen E-Commerce, April 2019
- Nr. 440: Serpil Taş, René Arnold:  
Auswirkungen von OTT-1-Diensten auf das Kommunikationsverhalten – Eine nachfrageseitige Betrachtung, Juni 2019
- Nr. 441: Serpil Taş, Christian Hildebrandt, René Arnold:  
Sprachassistenten in Deutschland (Juli 2019)



**ISSN 1865-8997**