

Andre Alpar, Claudius Konopka

# Beyond Touch – intelligente Assistenten: Mit Skills und Actions die nächste Generation von Plattformen schlauer machen

Intelligente Assistenten verbreiten sich weltweit in rasantem Tempo. Auf über 2,5 Milliarden Devices können Nutzer heute bereits mit Alexa, Siri, Google Assistant und Co. interagieren. Da sich das Anwendungsspektrum von Nutzer zu Nutzer unterscheidet, gilt es, die Plattformen den eigenen Bedürfnissen anzupassen und entsprechend über Third-Party-Applikationen zu erweitern, ganz ähnlich, wie die Mobile-Plattformen Android oder iOS mit Apps erweitert werden. Und diese Parallele lässt bereits erahnen, welch riesiges Potenzial in den Applikationen für intelligente Assistenten steckt. Wir stehen heute ganz am Anfang.

## Einordnung von Skills, Actions, Capsules und Shortcuts

Nachdem im ersten Teil der dreiteiligen Serie „Beyond Touch – intelligente Assistenten“ die Plattformen Alexa, Siri, Google Assistant und Co. detailliert beleuchtet wurden, gilt es in diesem zweiten Teil, den Blick auf die Erweiterungsmöglichkeiten dieser Plattformen zu richten. Denn erst über Applikationen, die bei Amazon und Microsoft als Skills bezeichnet werden, bei Google als Actions oder auch Actions on Google, bei Samsung als Bixby Capsules oder bei Apple als Siri Shortcuts, entfalten intelligente Assistenten ihr volles Potenzial, indem das native Anwendungsspektrum um die Services von Drittanbietern erweitert wird und so beispielsweise Informationen, Shops oder Angebote direkt angebunden werden. Im Grunde eine immer wiederkehrende technologische Grundidee. Sind es auf dem Desktop die Plattformen Windows oder macOS, die mit Programmen erweitert werden, auf dem Smartphone Android und iOS, die mit Mobile Apps erweitert werden, handelt es sich bei intelligenten Assistenten um Alexa, Bixby oder Google Assistant, die wiederum mit Skills, Capsules oder Actions (kurz: Sprachapplikationen) erweitert werden. Gelistet werden Sprachapplikationen auf den Webseiten der Plattform-Anbieter.

### DER AUTOR



**Andre Alpar** ist seit über 20 Jahren unternehmerisch im Bereich Online-Marketing tätig. Heutzutage ist er als Investor, Fachautor, Keynote Speaker und Podcaster (#askOMR) aktiv.

### DER AUTOR



**Claudius Konopka** ist Gründer und Geschäftsführer der Beyto GmbH, einer auf Voice und Conversational AI spezialisierten Beratungsagentur. Mit Beyto berät er Unternehmen zum Thema Voice und virtuelle Assistenten.



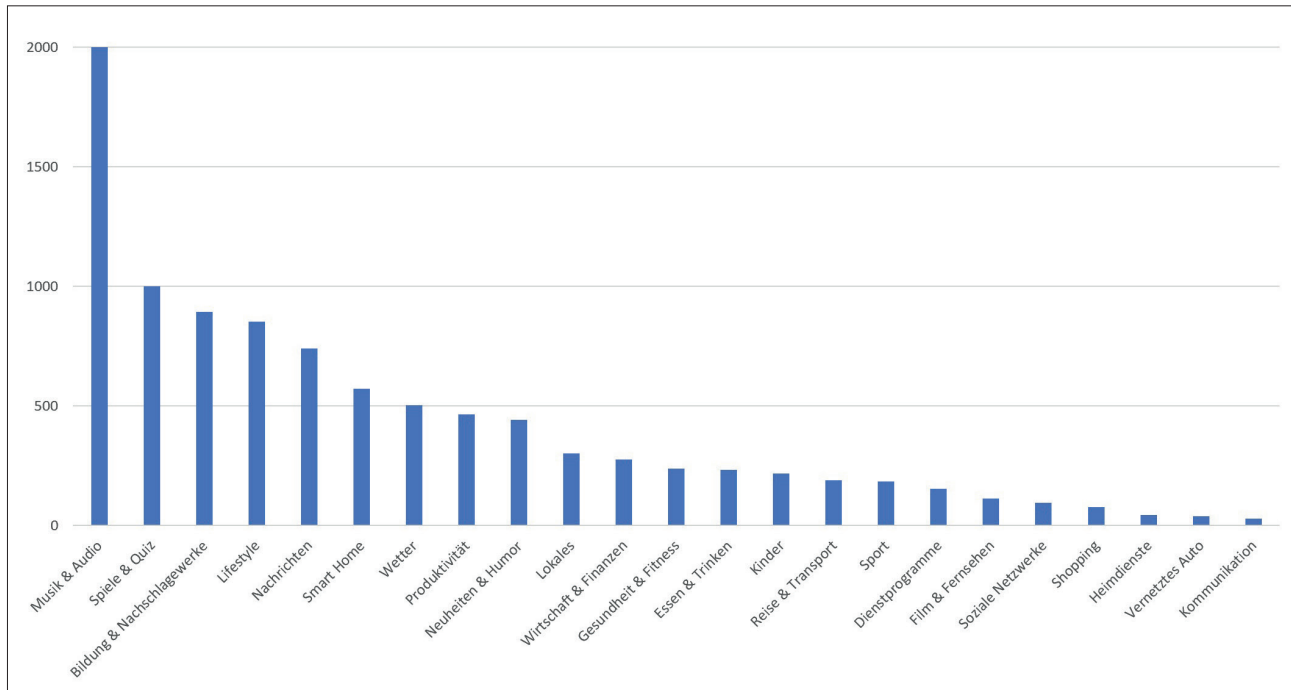


Abb. 1: Anzahl Amazon Alexa Skills je Kategorie (Stand Juni 2019)

Richtige App-Stores, wie es der Nutzer vom Apple App-Store oder Google Play-Store kennt, gibt es derzeit noch nicht.

## Was intelligente Assistenten nativ können und wofür es Applikationen braucht

Intelligente Assistenten beziehen ihre Informationen aus verschiedenen Quellen, allen voran aus Featured Snippets, Answer Boxes oder Knowledge Graphen der großen Suchmaschinen um Google, Bing oder auch Wolfram Alpha. Als weitere Quellen dienen zudem Wikipedia oder Pinterest. Dass bedeutet vereinfacht gesagt, dass alles, was Google weiß, spricht die gesamte Such- und Datenkompetenz, auch dem Google Assistant, Bixby oder Siri nativ zur Verfügung steht. Gleiches gilt für Bing in Bezug auf die Plattformen Alexa und Cortana. Alle Daten darüber hinaus müssen durch Third-Party-Applikationen bereitgestellt werden. Hier sind Amazon und Google derzeit führend und die einzigen Anbieter, deren Plattformen heute bereits mit Applikationen erweitert werden können. Für Microsofts Cortana sind Skills bisher lediglich in den USA verfügbar, wobei bezweifelt werden darf, dass Cortana

Skills in der jetzigen Form weitergeführt werden. Bei Samsung stehen die Bixby Capsules sowie der Marketplace gerade erst in den Startlöchern und Apple setzt aktuell auf Siri-Kurzbefehle (Shortcuts), welche mehr einer Bedienungshilfe als einer richtigen Sprachapplikation gleichen. Das alles zeichnet ein Bild, das verdeutlicht, wie sehr das Thema Sprachapplikationen noch in den frühen Anfängen steckt. Dies wird umso deutlicher mit Blick auf die Inhalte, die derzeit über Alexa Skills und Google Actions abgedeckt werden.

## Aktuelle Lage: Skills und Actions, die derzeit dominieren

Wirft man heute einen Blick auf die von Amazon und Google gelisteten Skills bzw. Actions, fühlt man sich zehn Jahre zurückversetzt, in die ersten Monate des Apple App-Stores. Damals wie heute sind die Applikationen der ersten Stunde geprägt von Spielereien, meist sehr simplen bis hin zu völlig sinnfreien Applikationen. Waren es im Jahr 2008 Anwendungen, die Furgeläusche nachmachten, Luftpolsterfolie imitierten oder den Bildschirm des Smartphones in ein Bierglas verwandel-

ten, wodurch der Betrachter sich durch hin und her schwappenden Schaum in eine Art Trancezustand versetzen konnte, sind es auch heute erneut Applikationen wie der „Pups-Generator“, „Tiergeräusche“ oder „Rülpfen“, die sich großer Beliebtheit erfreuen und die meisten Nutzerbewertungen auf sich vereinen. Die Verteilung der verfügbaren Skills verdeutlicht dies. So sind derzeit knapp 10.000 Skills (Stand Juni 2019) auf der Amazon-Website gelistet, von denen die Hälfte der Kategorie „Unterhaltung & Entertainment“ zuzuordnen ist. Zu den Top-Kategorien gehören aktuell „Musik & Audio“ mit circa 2.000 gelisteten Skills, „Spiele & Quiz“ mit etwa 1.000 geführten Skills oder „Lifestyle“ mit etwas weniger als 900 Skills. Unter „Smart Home“ werden nicht ganz 600 Applikationen geführt. Demgegenüber stehen 78 Skills in der Kategorie „Shopping“, was nicht einmal 1 % aller Skills bedeutet, oder etwa 30 Skills in der Kategorie „Kommunikation“. Ein ähnliches Bild, wenn auch in kleineren Zahlen, zeichnet sich auf der von Google bereitgestellten Seite für Actions on Google. Dass das Thema Sprachapplikationen derzeit noch in den Kinderschuhen steckt,

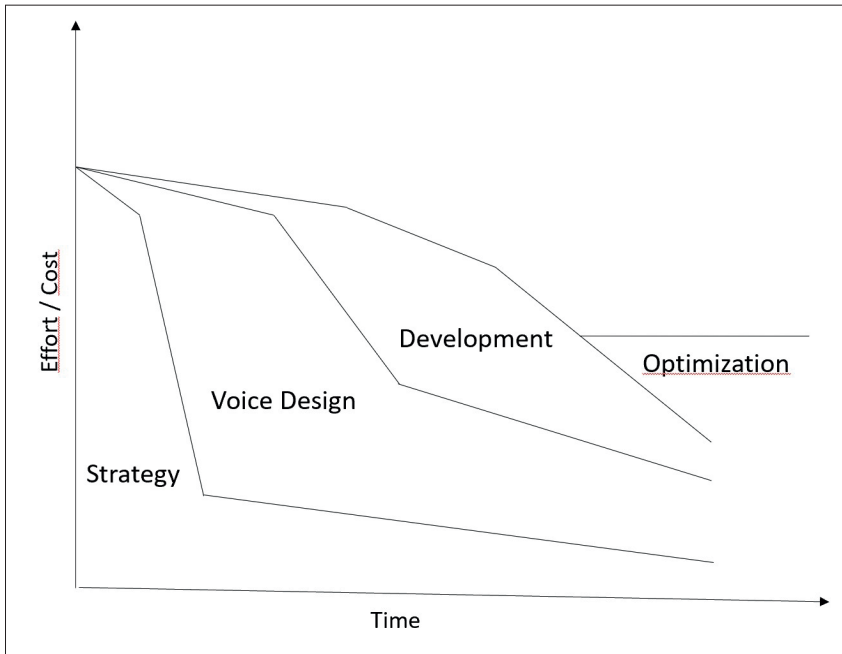


Abb. 2: Kosten-/Zeit-Einordnung für den Prozess zur Entwicklung einer Sprachapplikation

zeigen auch die absoluten Zahlen. Für Amazon Alexa sind derzeit weniger als 100.000 Applikationen weltweit verfügbar, für den Google Assistant noch nicht einmal ein Zehntel dessen. Das sind verschwindend geringe Zahlen. Wie damals auch schon, als die Menschen sich fragten, wofür sie Apps auf dem Smartphone überhaupt brauchen.

### Ausblick Entwicklung der Skill-/Actions-Stores

Dass die derzeitigen Anwendungen mit eher seichem Inhalt alsbald in den Hintergrund rücken werden, da Applikationen mit ernst zu nehmenden Use Cases nachrücken und in den Vordergrund drängen, davon ist auszugehen. Dass dies auch in kürzerer Zeit geschehen wird, als es seinerzeit beim Smartphone der Fall war, auch das ist anzunehmen. Allein durch die rasant wachsende Verbreitung intelligenter Assistenten, die eine deutlich schnellere Marktdurchdringung gegenüber dem Smartphone aufweisen, derzeit sind intelligente Assistenten auf circa 2,5 Milliarden Devices aufrufbar, steigen naturgemäß das Interesse und die Notwendigkeit von Drittanbietern, über diese Plattformen mit Nutzern in Kontakt zu treten. Auch führen neue techni-

sche Möglichkeiten bzw. Features, wie beispielsweise In-Skill-Purchases, zu völlig neuen Möglichkeiten der Monetarisierung und damit zwangsläufig auch zur nächsten Generation ernst zunehmender Applikationen mit echten Use Cases oder qualitativen Inhalten. Für mehr Inhalte in den Skill-Stores bringen die Plattformanbieter zudem auch stetig neue technische Lösungen, die das Erstellen sehr simpler Skills deutlich vereinfachen. Beispielsweise brachte Amazon Anfang des Jahres mit den Alexa Skill Blueprints eine Vielzahl von Templates heraus, die es jedem ermöglichen, eigene Applikationen ohne Programmierkenntnisse zu erstellen. Laut Amazon sind dadurch weltweit circa zwei Millionen neue Skills von Hunderttausenden Nutzern entstanden, die im privaten Umfeld millionenfach zum Einsatz kommen; bspw. Steuerung im Smart-Home-Bereich (<http://einfach.st/geekw2>). Wie sehr den Plattformanbietern daran gelegen ist, dass eigene Ökosystem zu pushen, zeigt die Tatsache, dass Amazon Entwickler von sehr beliebten Skills über ein Reward-Programm mit monatlich bis zu mittleren vierstelligen Beträgen incentiviert. Auch am Beispiel von Amazons Aktivitäten in Indien wird deutlich,

mit wie viel Druck in puncto Skill-Entwicklung agiert wird. Denn mittlerweile hat Amazon in Indien eine riesige Base von 40.000 Voice Developern aufgestellt, die über das gesamte Land hinweg Alexa Skills mit lokalem Touch entwickeln (<http://einfach.st/tech4>). Indien ist nach China die zweitstärkste Wirtschaftsmacht im asiatischen Raum und das vierte Land nach den USA, Großbritannien und Deutschland, in dem Amazon seinen Sprachassistenten eingeführt.

### Jetzt Handeln: Eigene Skills und Actions testen und lernen

An erster Stelle sollten Unternehmen sich vom derzeitigen Stand der Technologie virtueller Assistenten sowie vom vorhandenen Hardware-Fokus (Smart Speaker) wie auch von den vorherrschenden Unterhaltungs-Applikationen nicht beirren lassen. Virtuelle Assistenten stellen die nächste große Entwicklungsstufe in der Mensch-Maschinen-Kommunikation dar. Nach dem Desktop, dem Internet, dem Smartphone, bei denen der Nutzer stets mit der Maschine spricht, folgen nun intelligente Assistenten und erstmalig beginnt die Maschine, mit dem Menschen zu sprechen. Dass das derzeit noch alles etwas ruckelt, der Mensch der Maschine erst einmal das Sprechen beibringen muss, ist Teil dieser Evolution. Dass die Anbieter der Plattformen mit maximalen Ambitionen auf das Thema setzen, ist ersichtlich. Und dass intelligente Assistenten sehr schnell intelligenter werden, schneller, als heute vielleicht einiges darauf hindeutet, ist durchaus anzunehmen. Daher

empfiehlt es sich für Unternehmen, bereits jetzt das Thema Sprachtechnologie als Teil der eigenen Agenda zu sehen, die ersten Use Cases zu extrahieren, Sprachmodelle zu entwickeln und mit der technologischen Entwicklung im Gleichschritt zu wachsen. Frühzeitiges Experimentieren, schmale Lösungen und modulare Integrationen ermöglichen eine anhaltend hohe Lernkurve. Flexibilität und Schnelligkeit sind als Erfolgsfaktoren anzusehen. Denn die optimale akustische Kundeninteraktion beginnt sich gerade erst zu entwickeln! Praktisch täglich erfolgen Updates von Amazon, Google, Apple und Co.

### Beispielhafter Prozess zur Entwicklung einer Sprachapplikation

Der Prozess zur Entwicklung einer Sprachapplikation lässt sich in zehn aufeinanderfolgende Schritte gliedern, welche wiederum in die vier großen Kategorien Strategie, Voice Design, Development und Optimierung gruppiert werden. Den größten Zeitaufwand beansprucht dabei der Bereich Voice Design. Nachdem User Personas und Use Cases definiert, technische Möglichkeiten und Limitierungen sowie Format und Qualität vorliegender Daten mit den Zielen der eigenen Applikation synchronisiert wurden, beginnt die Design-Phase, in welcher die System Persona herausgearbeitet wird, sprich der Sound einer Marke, das Interaction Model kreiert und über Testing und Iteration auf Robustheit geprüft wird. Ziel ist es, ein intuitiv zu bedienendes Sprachmodell zu gewährleisten. Dafür ist von elementarer Bedeutung, die

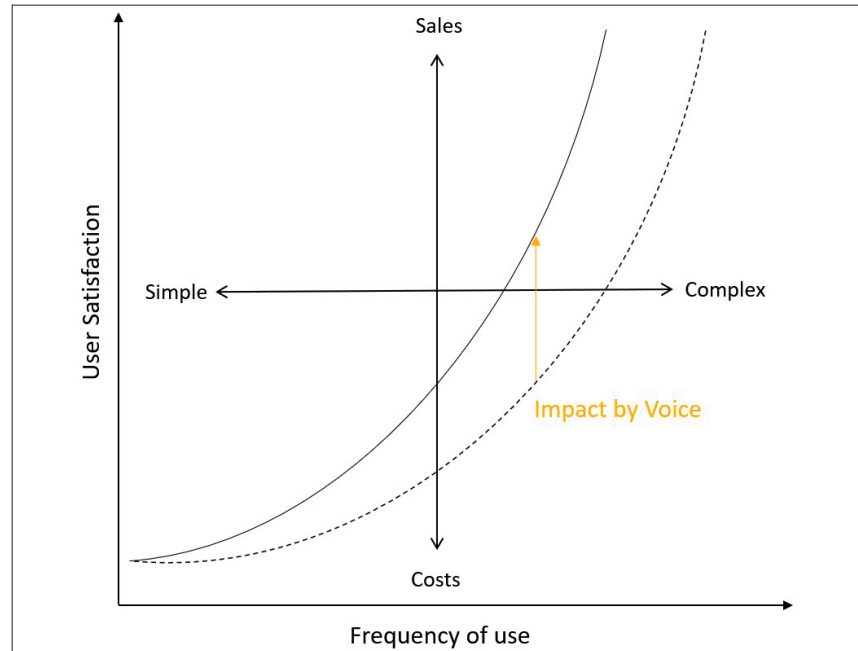


Abb. 3: Vereinfachte Darstellung zur Extrahierung des idealen Use Cases für Sprachsteuerung

Absichten des Nutzers korrekt vorherzusehen, insbesondere, wie dieser seine Absichten zum Ausdruck bringt, die richtigen Vorhersagen zu Phrasen und Redewendungen zu treffen. Die Sprachmodellierung ist von zentraler Bedeutung. Hat sich das Interaction Model bewährt, folgt die Skalierungsphase, in der der Long Tail ausgestaltet sowie das Design etwa multimodal erweitert wird. Sodann folgt die fortlaufende Analyse des Modells im Live-Betrieb sowie entsprechende Optimierungen. Eine Sprachapplikation lernt praktisch niemals aus und ist ein fortlaufender Prozess.

### Der richtige Use Case: Wann eine Sprachapplikation sinnvoll ist

Die Interaktion zwischen Mensch und Maschine über Sprache hat in vielerlei Hinsicht ganz entscheidende Vorteile. Sie ist deutlich schneller, bequemer wie auch barrierefreier. Sprache muss nicht neu gelernt werden, wie einst der einfache Klick auf den Button einer Website gegenüber dem Doppelklick zum Öffnen eines Programms, was selbst heute noch nicht jeder Nutzer verinnerlicht hat. Sprache, das ist die natürlichste Form der Kommunikation.

Und doch eignet sich die Spracheingabe nicht für jeden Use Case bzw. für jeden Kontext, in dem sich ein Nutzer befinden kann.

Daher stellt sich an erster Stelle die Frage, ob es für einen Nutzer komfortabel ist, sein Anliegen über Spracheingabe zu platzieren. Ist er imstande, über das Thema zu sprechen, kennt etwaige Fachtermini? Erfordert der Prozess der Eingabe einen kurzen oder ausschweifenden Dialog? Kann die Aufgabe via Multitasking erfolgen? Bedarf es der Hände und Augen des Nutzers? Das alles sind grundlegende Fragen, die darüber entscheiden, ob eine Sprachapplikation sinnvoll ist oder eben nicht. Am Ende sollte für jeden individuellen Case die Frage danach, ob Sprache gegenüber der Eingabe für das Keyboard überlegen ist, darüber entscheiden, ob es sich um den richtigen Use Case für eine Sprachapplikation handelt. ¶

*Im dritten und letzten Teil der Serie „Beyond Touch – intelligente Assistenten“ erwartet Sie in der nächsten Ausgabe ein Ausblick zum Horizont und auf die Attribute, über die der perfekte intelligente Assistent verfügen könnte, damit er sich nahtlos in unser Leben einfügt.*