

Sprich mit mir!

VOR- UND NACHTEILE SPRACHGESTEUERTER GERÄTE FÜR KINDER UND ERWACHSENE

David Kleeman

Der Autor erläutert die Vor- und Nachteile der Sprachsteuerung als Navigationsmedium von Smart Speakern, Handys, Spielen, Spielzeugen und Erlebnissen im Metaversum.

Wir sind uns immer noch nicht sicher, was »Sally« war, aber sie stand plötzlich auf unserer von Alexa generierten Einkaufsliste. Eindeutig hatte eine*r von uns die Sprachsteuerung von Alexa genutzt, um einen Artikel auf unsere Supermarkt-Einkaufsliste zu setzen, aber »sie« hatte sich verhöhrt. Es war ein guter Weckruf, dass Smart Speaker nur so smart bzw. schlau sind, wie es ihre Fähigkeit, uns Menschen zu verstehen, zulässt. Das kann besonders bei Kleinkindstimmen eine große Herausforderung darstellen. Die Ergebnisse können erschreckend sein ...

STEIGENDE NUTZUNG VON SPRACHGESTEUERTEN GERÄTEN BEI KINDERN

Die Verfügbarkeit von sprachgesteuerten Geräten in Haushalten steigt weltweit. In 16 von Dubit Trends¹ untersuchten Ländern hatten im Oktober 2021 46 % der 2- bis 15-jährigen Kinder Zugang zu einem Smart Speaker, ein Drittel davon besaß ein eigenes Gerät. In den USA liegen die Zahlen hinsichtlich Zugang und Besitz bei fast 56 %, fast eine Verdopplung im Vergleich zu 2018.

Wahrscheinlich sind die tatsächlichen Zahlen sogar noch höher, da es neben Smart Speakern noch viele weitere sprachgesteuerte Geräte gibt, u. a.

Smartphones. Mehr als 50 % der 2- bis 15-jährigen in den USA besitzen heutzutage ein eigenes Smartphone, in den anderen Ländern der Untersuchung liegt der Wert bei 43 %.

59 % der Kinder in den USA geben an, dass sie oft oder manchmal die Sprachsteuerungsfunktion nutzen, um zu finden, was sie anschauen, spielen oder anhören möchten. Im Oktober 2020 lag der Wert noch bei 50 %. Am häufigsten wird die Sprachsteuerung bei eigenständigen Smart Speakern genutzt (35 %), gefolgt von Smartphone (29 %) und TV-Fernbedienung (27 %).

KINDER UND SPRACHERKENNUNG: VOR- UND NACHTEILE

Spracherkennung hat eindeutig einige Vorteile für Kinder, besonders für die, die noch nicht lesen und schreiben

können und deshalb Probleme mit textbasierten Suchfunktionen oder Spielen haben, oder diejenigen, deren Feinmotorik für Touchscreens noch nicht ausreichend entwickelt ist. Gleichzeitig hat Sprache als Verbindungsmedium auch ihre Grenzen. Smart Speaker geben im Vergleich zu bildschirmbasierten Medien weniger hilfreiche Rückmeldungen, wenn die Anfrage eines jungen Nutzers/einer jungen Nutzerin keinen Treffer erzielt. Das Gerät wird dann entweder antworten, dass es nichts findet, was der Suchanfrage entspricht, oder es folgt eine Antwort basierend auf dem, was es »dachte«, gehört zu haben. Beispielsweise führt bei unserer Alexa zu Hause die Suche nach einem bestimmten Streaming-Nachrichtendienst regelmäßig zur Playlist einer Band mit einem annähernd ähnlichen Namen. Ist ein Bildschirm vorhanden (mit Text

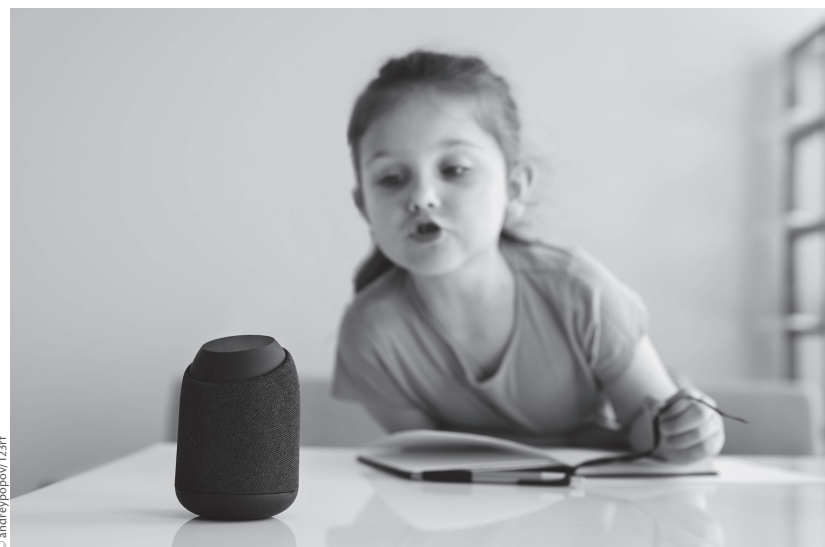


Abb. 1: Sprachgesteuerte Geräte werden in Haushalten weltweit immer häufiger eingesetzt

oder Sprachsteuerung), wird zusätzlicher Kontext bereitgestellt. Sage ich beispielsweise zu Alexa »Peppa Wutz«, antwortet sie: »Es klingt, als würdest du dich nach Peppa Wutz erkundigen. Würdest du gerne mehr darüber erfahren?« Eine bejahende Antwort ergibt einen lexikonartigen Vortrag zum Inhalt und den Synchronsprecher*innen der Serie – was für ein 3-jähriges Kind nicht besonders hilfreich ist. Im Gegensatz dazu führt die Google-Suche zu »Peppa Wutz« zu Bildern (Produkte und Screenshots), Links zu Videos, Spielen und Webseiten, Beschreibungen und vielem mehr. Das mag für ein Kleinkind überfordernd sein, aber zumindest bietet die Suche mehrere Anhaltspunkte, auf Grundlage derer die weitere Suche eingegrenzt werden kann.

Spracherkennung mittels künstlicher Intelligenz (KI) kann auch für Nicht-Kinderstimmen eine Herausforderung sein. Kürzlich wurde in einer Studie in den USA und Indien (Ramesh et al., 2022; zusammengefasst in Simonite, 2022) festgestellt, dass 40 % von mehr als 7.000 YouTube-Videos beliebter Kinderkanäle »Tabu«-Wörter in den Untertiteln enthielten, die automatisch aus den Soundtracks der Programme generiert wurden.

Spracherkennung ist ein noch junges Phänomen, das parallel zur Weiterentwicklung der künstlichen Intelligenz reift. Je mehr Merkmale die Programmierer*innen aus der realen Welt analysieren – ein breites Spektrum an Stimmen, Akzenten, Schlüsselwörtern, Satzbau und Nebenfaktoren etc. –, desto besser wird die KI Suchen schließlich ausführen und ihre umfangreichen internen Wissensbestände erweitern, um schneller herauszufinden, was ein Kind möchte.

STIMME IM METAVERSUM

Auf unserem Weg zum Metaversum (Kleeman, 2021) ist die Spracherkennung von großem Interesse. In

immersiven Welten kann vieles mit Hand-Controllern gesteuert werden und die Verwendung von Text zur Navigation komplexer Aktivitäten oder für die Bewegung von einem Ort zum anderen würde die Illusion und die »suspension of disbelief« (»Aussetzung der Ungläubigkeit«) zerstören. Diese Welten werden für soziales Spiel erschaffen und deshalb möchten wir mit anderen mittels Sprache kommunizieren.

Im Metaversum werden wir mehr als nur schöne Worte brauchen, um den Kontext herzustellen, der derzeit noch fehlt. In der virtuellen Realität könnte die sprachaktivierte KI beispielsweise das Gehörte durch Eye-Tracking (Blickverfolgung), Körperpositionierung und Interferenzen aus der Umgebung ergänzen. Meta hat vor Kurzem eine Technologie namens »BuilderBot« vorgestellt, die Gegenstände in einer virtuellen Umgebung auf der Grundlage von gesprochenen Wörtern der Nutzer*innen und kontextuellen Hinweisen im virtuellen Raum platziert (Ghaffary, 2022).

Wir müssen sicherstellen, dass das Metaversum ein integrativer und allen zugänglicher Raum ist, und Sprachsteuerung wird eine wichtige Rolle dabei spielen, beispielsweise Menschen mit und ohne Behinderung eine gute Erfahrung zu ermöglichen. Sprache kann das Navigationsmedium für jene sein, deren Bewegungsmöglichkeiten eingeschränkt sind und die deswegen keinen Hand-Controller nutzen oder ihre echten Körper in virtuellen Räumen bewegen können.

SPRICH MIT MIR!

Erinnern Sie sich noch an die Zeit, als die ersten Tablets auf den Markt kamen und die Kinder dachten, die ganze Welt bestünde aus interaktiven Touchscreens? YouTube war voll mit Videos, in denen Kleinkinder versuchten, Bücher und Zeitschriften zu »swipen« und anzutippen.

Vielleicht werden kleine Kinder bald denken, dass sie alles mit ihrer Sprache steuern können. Und mit immer mehr sprachgesteuerten Geräten – Smart Speakern, Smartphones, Smart Homes, Smart-Fernbedienungen und (vermutlich auch smarten) Spielzeugen – liegen sie damit vielleicht gar nicht so falsch? ■

ANMERKUNG

¹ Dubit Trends wird jedes Jahr im April und Oktober in den USA (2.000 Kinder/Eltern 2-18) und im Vereinigten Königreich (1.300 Kinder/Eltern 2-18) durchgeführt. 18 weitere Länder werden im Wechsel befragt (1.000 Kinder/Familien 2-16). Es handelt sich um eine Online-Umfrage, die repräsentativ für das nationale Bevölkerungsprofil ist und die ethnische Zugehörigkeit, das Haushaltseinkommen, das Alter der Eltern und mehr erfasst. Die Studie wird an Abonnent*innen auf der ganzen Welt weitergegeben und dient als Grundlage für die Forschung und Strategieberatung von Dubit sowie für die Entwicklung und Produktion von Spielen, Apps und Erlebnissen im Metaversum.

LITERATUR

Ghaffary, Shirin (2022). Mark Zuckerberg wants to build a voice assistant that blows Alexa and Siri away. *Vox recode* (February 23, 2022). Verfügbar unter: <https://www.vox.com/recode/22948097/mark-zuckerberg-voice-assistant-metaverse-ai-announcement> [28.4.22]

Kleeman, David (2021). What is the »metaverse« & why does it matter? Verfügbar unter: <https://staging.kidscreen.com/2021/09/14/what-is-the-metaverse-why-does-it-matter/> [28.4.22]

Ramesh, Krithika, KhudaBuksh, Ashiqur & Kumar, Sumeet (2022). »Beach« to »bitch«: Inadvertent unsafe transcription of kids' content on YouTube. *Preprint* (February 17, 2022). Verfügbar unter: https://github.com/sumeetkr/UnsafeTranscriptionofKidsContent/blob/main/YouTube_Transcription_Final.pdf [28.4.22]

Simonite, Tom (2022). YouTube's captions insert explicit language in kids' videos. *Wired* (February 24, 2022). Verfügbar unter: <https://www.wired.com/story/youtubes-captions-insert-explicit-language-kids-videos/> [28.4.22]

DER AUTOR

David Kleeman ist Vizepräsident Global Trends von Dubit, einer Strategie- und Forschungsberatung mit Sitz in Leeds, England.

