

# Sprechende Maschinen

**KÜNSTLICHE INTELLIGENZ** Synthetische Stimmen entwickeln sich gerade rasant weiter. Immer mehr Roboter, Geräte und Systeme können sprechen. Aber will man, dass das Auto spricht? Oder der Lift? Und sollen sie menschenähnlich klingen? Sprachsynthese wirft etliche Fragen auf.

VON OLIVER BENDEL

Schon vor tausenden Jahren träumte der Mensch davon, künstliche Kreaturen zu schaffen. Diese sollten seine Gehilfen sein, seine Begleiter, Geliebten und Krieger. Bei Homer und Ovid lernen wir sie kennen, schätzen und fürchten. Die meisten dieser Kreaturen, auch in Mittelalter, Renaissance und Barock, waren stumm. Damit wurde die Kluft zwischen ihnen und uns angedeutet. Eine Ausnahme bildeten die sprechenden Köpfe, die in die Zukunft blicken konnten. Heute tragen uns Text-to-Speech-Systeme die Inhalte von Websites vor, plaudern Sprachassistenten wie Siri und Alexa mit uns und melden sich Fahrzeuge laut und vernehmlich zu Wort. Ein weites Feld aus technischer, psychologischer und ethischer Perspektive.

## VERTRAUTE DIALOGSYSTEME

Einige Chatbots beherrschen nicht allein Text-, sondern auch Sprachausgabe. Sgt. Star von der US Army versucht die Amerikaner mit dunkler Stimme für seine Sache zu gewinnen. Sprachassistenten wie Siri und Google Assistant haben hellere Stimmen und ihre Stärken weniger im Militärischen als vielmehr im Privaten und Ökonomischen. Sie sagen uns, was «SSML» oder «DSGVO» bedeutet und wo sich das nächste vegetarische Restaurant befindet. Bei manchen Modellen könnten wir die Stimme wechseln. Aber wir behalten meist die Voreinstellungen bei, die eben mit weiblichen, jungen Charakteren verbunden sind.

Anbieter wie Amazon bemühen sich darum, dass die Stimmen immer menschlicher tönen. So kann Alexa seit einiger Zeit flüstern, dank einem Befehl der Speech Synthesis Markup Language (SSML). An Google Duplex, dem auf Google Assistant basierenden Projekt, ist einiges bemerkenswert. Mit einer High-End-Technologie erreichen wir fast jeden Ort der Welt. Wir könnten sie in unserem Namen einen Termin bei einem

Coiffeur in einem Dorf in Nordindien vereinbaren lassen, der nicht einmal merkt, dass es sich um eine Maschine handelt. Es ist zum einen die Stimme, die lebensecht wirkt, zum anderen die Sprechweise. Das System gerät ins Stocken, es streut Mmhs und Ähs ein. Wenn es sich nun noch räuspert und die Nase hochzieht, kann es niemand mehr von einem Menschen unterscheiden.

Schon Ovid schrieb über die Frauenstatue, die Pygmalion geformt hatte: «Dass es nur Kunst war, verdeckte die Kunst.» Aphrodite, in die der Bildhauer eigentlich verliebt war, erweckte sie aus Mitleid zum Leben. Die Kunst bei Google Duplex besteht in der Imperfektion. Aus dieser entsteht die Perfektion. Aber sollten Maschinen überhaupt lebensecht klingen? Sollten sie uns einreden, dass sie Menschen sind? Dies wurde aus ethischer Perspektive intensiv diskutiert. Wir haben bereits vor Jahren gefordert, dass ein Dialogsystem deutlich machen soll, dass es keine Person ist. Und 2013 eines präsentiert, das genau dies umsetzt, den GOODBOT, der zuerst in dieser Zeitung behandelt wurde. Auch Google Duplex räumt inzwischen am Anfang des Gesprächs ein, bloss eine Maschine zu sein. Man muss es nur hören und verstehen.

In einem Projekt mit einer deutschen Universität werden wir testen, wie künstliche Stimmen auf Probanden wirken. Ich habe ein Set produziert, das Sätze mittels der erwähnten Auszeichnungssprache für Sprachsynthese variiert. So wird eine Aussage einmal neutral, einmal mit Begeiste-

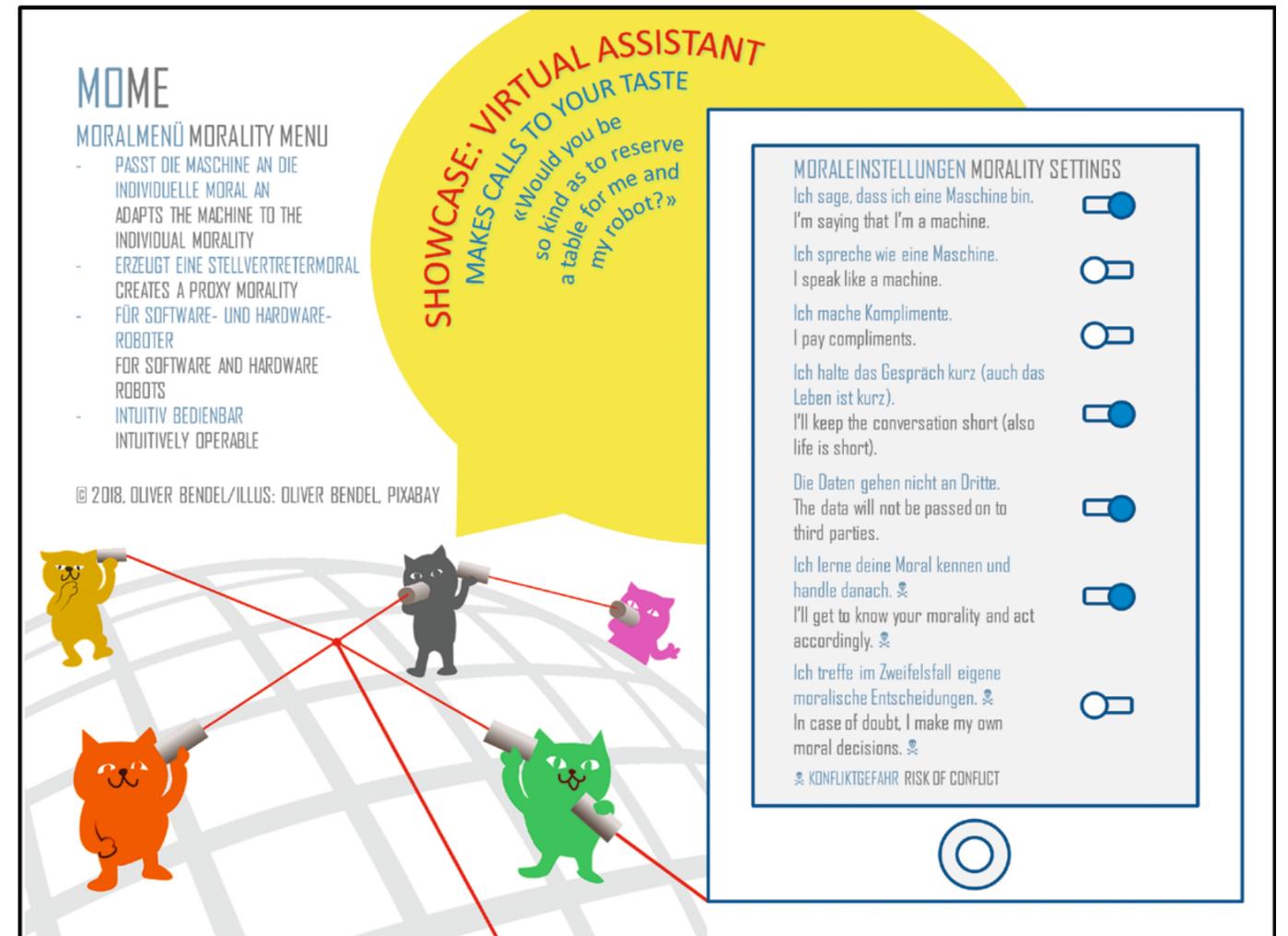
rung vorgetragen. In einem anderen Beispiel habe ich die Stimme jünger, in einem weiteren weicher gemacht. Mit SSML können wir Tonhöhe, Aussprache, Betonung usw. verändern. Ich bin daran interessiert, die Stimmen zu perfektionieren. Aber auch daran, dass die Gesprächspartner nicht getäuscht werden.

## PEPPER UND CO.

Roboter Pepper, eigentlich als Freund und Familienmitglied gedacht, ist immer häufiger in Einkaufszentren anzutreffen, in Kalifornien wie im Kanton Zürich, zudem in Alten- und Pflegeheimen. Der Hersteller hat uns erklärt, warum die Stimme kindlich und robotisch klingt. Zum einen passt sie damit zum Äusseren, zum anderen erzeugt sie eine gewisse Distanz. Er hat uns zudem erzählt, dass Heranwachsende, die oft mit Pepper allein sind, seine Bewegungen, seine Stimme und seine Sprechweise nachahmen. Für sie ist die Stimme nicht unbedingt fremdartig. Sie leben in einer anderen Welt. Wenn sich dagegen Pepper im Glattzentrum an uns Erwachsene wendet, fühlen wir sofort die Kluft zwischen ihm und uns. Er kann sprechen, anders als die früheren künstlichen Kreaturen, aber auf eine merkwürdige Weise. Im Alten- und Pflegeheim kann es sein, dass wir uns nicht ernstgenommen fühlen von ihm. Der Ersatz der Stimme führt nicht zwangsläufig zum Ziel, denn äusserlich bleibt er ein Kind.

## HEY MERCEDES

Man streitet darüber, ob interaktive auditive Systeme in Fahrzeugen etwas verloren haben. Ein Kollege von mir behauptet, dass sie ablenken. Das stimmt einerseits, andererseits haben wir schon immer mit unseren Sitznachbarn geschwätzt und das Mobiltelefon benutzt. 2015 habe ich in einem Artikel mitteilungs- und darstellungsfreudige Autos thematisiert, die sowohl eine Stimme als auch einen Charakter haben. Als ich von Olli hörte, dem smarten Shuttle, das sich



Das Moralmenü für Sprachassistenten (eigene Darstellung).

Grafik: zvg

mit Hilfe von IBM Watson mit den Insassen unterhält, war ich nicht überrascht. Er fragt die Insassen, ob sie ein Eis wollen, und wenn alle «Certo» schreien, fährt er sie zur Gelateria. Als Daimler mit seinem auditiven System herauskam, war ich vollends davon überzeugt, dass dies die Zukunft ist. Mit dem Kommando «Hey Mercedes» aktiviert man es. Es versteht Aussagen zur Bedienung der Unterhaltungsgeräte und des Fahrzeugs und lernt ständig dazu. Es stellt sich auf den Benutzer und dessen Stimme ein und hat nach kurzer Zeit kaum noch Mühe mit Dialekt und Akzent.

Hier wird ein weiteres Problem sprechender Maschinen offenbar. Indem sie nicht nur mit uns sprechen, sondern wir auch mit ihnen, öffnen wir uns ihnen gegenüber. Die Stimme verrät Geschlecht, Alter und Gesundheit, die Sprechweise den Bildungsgrad, die Sprache die Herkunft. Von Bedeutung ist, wie wir etwas sagen und was wir sagen. Siri, Alexa, Pepper und Mercedes belauschen uns, und die Datenschutz-Grundverordnung (DSGVO) ist für

sie ebenso relevant wie für Websites und E-Mails.

## DAS LETZTE WORT

Mit sprechenden Fahrzeugen werden wir uns abfinden, sprechende Fahrstühle sind uns vertraut. Aber wollen wir uns wirklich von jedem Ding von der Seite anquatschen lassen? Egal, wie man die Stimme gestaltet, egal, wie oft sie ausruft, dass sie einer Maschine gehört – das Sprechen zeichnet uns aus, und es mag anmassend erscheinen, wenn ein Kühlschrank zu reden beginnt. Neben solchen Akzeptanzproblemen ergeben sich Fragen aus Sicht der Ethik. Warum sind die Stimmen mehrheitlich weiblich, warum sanft und jung, wie ist damit umzugehen, dass die Maschinen eines Tages besser sprechen als wir? Dies reflektieren wir in Informations- und Roboterethik und überhaupt in der Roboterphilosophie. In der Maschinenethik reflektieren wir nicht bloss, sondern wir implementieren, wir kreieren Dialogsysteme, Chatbots und Sprachassistenten, die sich moralisch verhalten. Den

GOODBOT habe ich erwähnt. Für Lösungen wie Google Duplex habe ich im Mai 2018 ein Moralmenü erfunden, mit dem man sie an die Überzeugungen und Vorstellungen des Benutzers anpassen könnte. Das letzte Wort bei Dialogsystemen ist noch nicht gesprochen.

## DER AUTOR



Dr. Oliver Bendel ist studierter Philosoph und promovierter Wirtschaftsinformatiker. Er lehrt und forscht als Professor an der Hochschule für Wirtschaft FHNW mit den Schwerpunkten Wissensmanagement, Wirtschafts-, Informations- und Maschinenethik.

Foto: Kai R. Jochim