

Christoph Drösser

Wenn die Dinge mit uns reden

Von Sprachassistenten,
dichtenden Computern
und Social Bots



Dudenverlag

Christoph Drösser

Wenn die Dinge mit uns reden

Für Andrea

Christoph Drösser

Wenn die Dinge mit uns reden

Von Sprachassistenten,
dichtenden Computern und Social Bots

Dudenverlag
Berlin

INHALT

Vorwort	6
Die Dinge lernen sprechen	9
Der Traum von der sprechenden Maschine <i>Die Geschichte der Sprachtechnologie</i>	15
Geschwätzige Helfer <i>Die Welt der Chatbots</i>	31
Von der schnarrenden Roboterstimme zur realistischen Simulation <i>Wie man Maschinen zum Reden bringt</i>	50
Sprachassistenten für alle Gelegenheiten <i>Siri, Alexa und Co.</i>	65
Verstehen, ohne zu wissen <i>Die Möglichkeiten und Grenzen statistischer Sprachanalyse</i>	85
Roboterjournalisten und Fake-News-Generatoren <i>Der Computer als Autor</i>	106
Partner oder Werkzeug <i>Wie wir uns unsere Sprachassistenten wünschen</i>	133
Anhang	148

VORWORT

Wenn Sie dieses Buch lesen, ist es bereits veraltet. Tut mir leid, aber das ist unvermeidlich. Das Gebiet der Computerlinguistik (auf Englisch *Natural Language Processing*, NLP) gehört zu den dynamischsten Disziplinen der künstlichen Intelligenz (KI), und die Zyklen der Innovation sind kürzer als der Lebenszyklus eines gedruckten Buchs. Nur ein Beispiel: Ich habe mit Miles Brundage von OpenAI ein Interview über das Sprachmodell GPT-2 geführt (siehe Seite 116), und einen Tag später kündigte die Organisation den Nachfolger GPT-3 an: noch größer, noch verrückter, noch leistungsfähiger. Ich habe dann schnell noch einige Änderungen im entsprechenden Kapitel gemacht – aber wenn das Buch ausgeliefert wird, gibt es wahrscheinlich schon wieder eine neue Entwicklung auf diesem Gebiet, das mit erstaunlichen Leistungen auf sich aufmerksam macht.

Das Buch ist also eine Momentaufnahme des Stands der Technik im Sommer 2020. Aber die Buchform ermöglicht es auch, einen Schritt zurückzutreten und allgemeine Entwicklungen zu beschreiben, die unabhängig von einzelnen Produkten und Trends sind – das habe ich versucht. Wir befinden uns an einer neuen Schwelle der technischen Entwicklung, wir machen den Schritt in ein Zeitalter, in dem Maschinen den von Alan Turing 1950 vorgeschlagenen »Intelligenztest« bestehen und mit uns Konversationen in natürlicher Sprache führen. Wie diese Zu-

kunft gestaltet wird, das wird jetzt entschieden, in wissenschaftlichen Labors und in den Entwicklungsabteilungen der großen Technikkonzerne.

Die heiße Schreibphase des Buchs fiel zusammen mit dem Ausbruch der Corona-Pandemie. Eigentlich ideal für jemanden, der sich ohnehin zum einsamen Arbeiten zurückziehen will – aber da ist noch die Familie, die auch nicht mehr das Haus verlässt, um zur Arbeit oder zur Schule zu gehen. Ich danke meiner Frau Andrea Cross und meinem Sohn Oliver, dass wir es zusammen durch diese ungewöhnlichen Monate geschafft haben.

Danken möchte ich außerdem meiner Agentin Heike Wilhelmi und meinen Lektorinnen Juliane von Laffert und Silke Körber, die wertvolle Anregungen gegeben haben. Mein Freund und Kollege John Markoff, der seit über 40 Jahren die Entwicklung im Silicon Valley verfolgt, hat mir bei der Anbahnung einiger Kontakte in der Tech-Welt geholfen.

San Francisco, im Sommer 2020

Christoph Drösser

DIE DINGE LERNEN SPRECHEN

Reden Sie mit Ihrem Smartphone? Ich tue es seit einiger Zeit: »Hey Siri, stell einen Timer auf 20 Minuten!« – »Wie hoch ist der Mount Everest?« – »Überweise 100 Euro an Peter!« Etwas komisch komme ich mir dabei immer noch vor, aber man gewöhnt sich daran. Hätte ich einen Amazon-Echo- oder einen Google-Home-Lautsprecher im Haus, könnte ich auch Kommandos geben wie »Stell die Heizung auf 20 Grad« oder »Mach das Licht im Bad aus«, so wie es zwölf Prozent der Deutschen bereits tun. Die menschliche Stimme ist das neue Interface; mit ihr kann man viele Kommandos leichter und schneller geben als über eine Tastatur oder einen Bildschirm mit vier Untermenüs.

Das Wort »Kommando« beschreibt es allerdings schon recht treffend: Auf dem Exerzierplatz findet die Kommunikation meist in einer Richtung statt, vom Offizier zum Untergebenen. Es wird keine Antwort erwartet, sondern eine bestimmte Aktion. Von einem Gespräch kann nicht die Rede sein. So ist es im Moment auch noch mit unseren Geräten, aber in Zukunft wird die Maschine immer öfter antworten. Wir treten zum ersten Mal in der Menschheitsgeschichte in einen Dialog mit nichtmenschlichen Entitäten – »Wesen« möchte ich nicht sagen. Die Sprachfähigkeit der Dinge um uns herum hat sich in wenigen Jahren sprunghaft verbessert, und dieses Buch soll einen Überblick darüber geben, wie gut Maschinen unsere Sprache heute schon beherrschen.

Aber reden die Dinge wirklich schon mit uns, wie es der Titel dieses Buchs suggeriert? Sicherlich sollen sie den Eindruck erwecken – die schnarrenden Roboterstimmen sind abgelöst worden von sanften, täuschend echt klingenden Männer- oder (meistens) Frauenstimmen, die teilweise sogar »Ähs« und »Hms« in ihren Redefluss einbauen. Wir können gar nicht anders, als in ihrem Tonfall und ihrer Stimmelmelodie nach Anzeichen für eine Persönlichkeit zu suchen. Um diese Wirkung zu verstärken, haben Designer den Geräten einen ganzen Katalog an vorgefertigten Antworten mitgegeben, die sie witzig, ironisch oder sogar sexy erscheinen lassen. Allerdings handelt es sich derzeit noch um »auswendig gelernte« Phrasen, die uns Menschlichkeit suggerieren sollen. Und viele lassen sich gern davon verführen, auch wenn echte Liebesgeschichten zwischen Mensch und Maschine noch Stoff für Science-Fiction sind – wie in dem Film »Her« von 2013, in dem ein Nutzer sich in seine digitale Assistentin verliebt.

Dass sich die Sprachfähigkeit der Geräte um uns herum in den letzten Jahren so explosionsartig verbessert hat, ist vor allem auf Techniken zurückzuführen, die seit 2010 das gesamte Gebiet der künstlichen Intelligenz (KI) revolutioniert haben: neuronale Netze und Deep Learning. Solche Netzwerke erkennen in selbstfahrenden Autos die Verkehrslage, sie schlagen menschliche Gegner im japanischen Go-Spiel – und sie erwerben auch auf eine ganz neue Art sprachliche Fähigkeiten. Sie werden in diesem Buch des Öfteren auf Sätze wie diesen treffen: »Seit neuronale Netze dafür eingesetzt werden, hat sich die Leistung der Systeme sprunghaft verbessert.« Es gibt auf dem Gebiet der natürlichen Sprachverarbeitung (NLP, von *Natural Language Processing*, im Deutschen spricht man auch von Computerlinguistik), keine Unterdisziplin, die von dieser Technik nicht tangiert worden wäre.

Kurz gesagt, erlauben neuronale Netze es Maschinen zu ler-

nen, ähnlich wie ein Mensch lernt. Bei aller Euphorie ist es wichtig, das Wort »ähnlich« zu betonen – ein neuronaler Computer ist kein biologisches Gehirn, und die Netze stoßen zunehmend an Grenzen. Dennoch ist es faszinierend zu sehen, was diese auf sehr simplen Prinzipien beruhende Technik in wenigen Jahren geleistet hat und wahrscheinlich noch leisten wird.

Menschen lernen eine Sprache auf zwei sehr unterschiedliche Weisen. Beginnen wir mit der Zweit- oder Drittsprache in der Schule oder im Erwachsenenalter: Wenn wir uns eine Fremdsprache aneignen wollen, dann müssen wir Vokabeln lernen (also die Bedeutung der Wörter anhand ihrer Übersetzung in unsere Muttersprache) und grammatische Regeln verinnerlichen: Welche Endungen bekommen die Wörter je nach Fall oder Zeit (Deklination und Konjugation), in welcher Folge werden sie in einem korrekten Satz aneinandergereiht (Subjekt/Prädikat/Objekt)? Alle frühen Versuche, Computern Sprache beizubringen, haben sich an diese Vorgehensweise angelehnt. So musste – im Fall von gesprochener Sprache – der Strom von akustischen Signalen zunächst in einzelne Laute, »Phone« genannt, segmentiert werden, aus denen Wörter zusammengesetzt wurden. Letztere mussten dann auf ihre grammatische Funktion hin überprüft werden, um den Sinn des Gesprochenen zu analysieren.

Diese abstrakte, auch »symbolisch« genannte Methode der Sprachverarbeitung ist schwer in die maschinelle Praxis zu überführen, etwa in Übersetzungssysteme. Sie ist mit viel »Handarbeit« der Programmierer verbunden und hat in der Vergangenheit zu wenig befriedigenden Ergebnissen geführt. Die neuen Verfahren maschinellen Lernens orientieren sich weniger an abstrakten Grammatikregeln, sondern nehmen sich eher den Erwerb unserer Muttersprache zum Vorbild, die jedes Kind auf dem Globus nach maximal fünf Jahren fast perfekt beherrscht.