

Stefan Weber



TELEPOLIS

Roboterjournalismus, Chatbots & Co.

Wie Algorithmen Inhalte produzieren
und unser Denken beeinflussen

Mit einem Vorwort von Uwe Hannig und einem
juristischen Kommentar von Albrecht Haller



TELEPOLIS

→ www.telepolis.de

Das Online-Magazin TELEPOLIS wurde 1996 gegründet und begleitet seither die Entwicklung der Netzkultur in allen Facetten: Politik und Gesetzgebung, Zensur und Informationsfreiheit, Schutz der Privatsphäre, wissenschaftliche Innovationen, Entwicklungen digitaler Kultur

in Musik, Film, bildender Kunst und Literatur sind die Kernthemen des Online-Magazins, welche ihm eine treue Leserschaft verschafft haben. Doch TELEPOLIS hat auch immer schon über den Rand des Bildschirms hinausgesehen: Die Kreuzungspunkte zwischen realer und virtueller Welt, die »Globalisierung« und die Entwicklung der urbanen Kultur, Weltraum und Biotechnologie bilden einige der weiteren Themenfelder.

Als reines Online-Magazin ohne Druckausgabe nimmt TELEPOLIS damit eine einzigartige Stellung im deutschsprachigen Raum ein und bildet durch seine englischsprachige Ausgabe und seinen internationalen Autorenkreis eine wichtige Vermittlungsposition über sprachliche, geografische und kulturelle Grenzen hinweg. Verantwortlich für das Online-Magazin und Herausgeber der TELEPOLIS-Buchreihe ist Florian Rötzer.

Die TELEPOLIS-Bücher basieren auf dem Themenkreis des Online-Magazins. Die Reihe schaut wie das Online-Magazin über den Tellerrand eingefahrener Abgrenzungen hinaus und erörtert Phänomene der digitalen Kultur und der Wissensgesellschaft.

Eine Auswahl der bisher erschienenen
TELEPOLIS-Bücher:

Stefan Weber

Das Google-Copy-Paste-Syndrom

Wie Netzplagiate Ausbildung und

Wissen gefährden

2008, 182 Seiten, 16,00 €

Matthias Brake

Mobilität im regenerativen Zeitalter

Was bewegt uns nach dem Öl?

2009, 154 Seiten, 16,00 €

Stefan Selke / Ullrich Dittler

Postmediale Wirklichkeiten

Wie Zukunftsmedien die

Gesellschaft verändern

2009, 256 Seiten, 19,00 €

Lothar Lochmaier

Die Bank sind wir

Chancen und Perspektiven von

Social Banking

2010, 160 Seiten, 15,90 €

Matthias Becker

Datenschatten

Auf dem Weg in die

Überwachungsgesellschaft?

2010, 184 Seiten, 16,90 €

Harald Zaun

**SETI – Die wissenschaftliche Suche
nach außerirdischen Zivilisationen**

Chancen, Perspektiven, Risiken

2010, 320 Seiten, 19,90 €

Marcus B. Klöckner

9/11 – Der Kampf um die Wahrheit

2011, 218 Seiten, 16,90 €

Hans-Arthur Marsiske

**Kriegsmaschinen –
Roboter im Militäreinsatz**

2012, 252 Seiten, 18,90 €

Nora S. Stampfl

Die verspielte Gesellschaft

Gamification oder Leben im Zeitalter des

Computerspiels

2012, 128 Seiten, 14,90 €

Nora S. Stampfl

Die berechnete Welt

Leben unter dem Einfluss von Algorithmen

2013, 124 Seiten, 14,95 €

Christian J. Meier

**Eine kurze Geschichte des Quanten-
computers**

Wie bizarre Quantenphysik eine neue

Technologie erschafft

2015, 188 Seiten, 16,90 €

Michael Firmkes

Das gekaufte Web

Wie wir online manipuliert werden

2015, 324 Seiten, 18,95 €

Klaus Schmeh

Versteckte Botschaften

Die faszinierende Geschichte

der Steganografie

2017, 318 Seiten, 19,95 €

Christian J. Meier

Suppenintelligenz

Die Rechenpower aus der Natur

2017, 246 Seiten, 16,90 €

Weitere Informationen zu den TELEPOLIS-Büchern und Bestellung unter:

→ www.dpunkt.de/telepolis

TELEPOLIS





Doz. Dr. Stefan Weber, geboren 1970 in Salzburg. Universitätslektor an der Universität Wien und Senior Researcher der Research Studios Austria. Bekannt wurde er als Plagiatsgutachter und Verfasser kulturkritischer Bücher zum Wandel der Textkultur durch die Digitalisierung (»Das Google-Copy-Paste-Syndrom«, 2. Auflage dpunkt/Telepolis 2008, »Die Medialisierungsfalle«, 2008). Er bloggt unter <https://plagiatsgutachten.de/blog>.

Stefan Weber

Roboterjournalismus, Chatbots & Co.

Wie Algorithmen Inhalte produzieren und unser Denken beeinflussen

Mit einem Vorwort von Uwe Hannig und
einem juristischen Kommentar von Albrecht Haller



Heise

Stefan Weber, weber@plagiatsgutachten.de

Reihenherausgeber: Florian Rötzer, München, fr@heise.de

Lektorat: Christian Reifert

Copy-Editing: Susanne Rudi, Heidelberg

Satz: Veronika Schnabel

Herstellung: Stefanie Weidner

Umschlaggestaltung: Hannes Fuß, www.hannesfuss.de

Druck und Bindung: M.P. Media-Print Informationstechnologie GmbH, 33100 Paderborn

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN:

Print 978-3-95788-104-5

PDF 978-3-95788-986-7

ePub 978-3-95788-987-4

mobi 978-3-95788-988-1

1. Auflage 2019

Copyright © 2019 Heise Gruppe GmbH & Co. KG

Karl-Wiechert-Allee 10

30625 Hannover

Die vorliegende Publikation ist urheberrechtlich geschützt. Alle Rechte vorbehalten. Die Verwendung der Texte und Abbildungen, auch auszugsweise, ist ohne die schriftliche Zustimmung des Verlags urheberrechtswidrig und daher strafbar. Dies gilt insbesondere für die Vervielfältigung, Übersetzung oder die Verwendung in elektronischen Systemen. Alle Informationen in diesem Buch wurden mit größter Sorgfalt kontrolliert. Weder Herausgeber, Autor noch Verlag können jedoch für Schäden haftbar gemacht werden, die in Zusammenhang mit der Verwendung dieses Buches stehen.

5 4 3 2 1 0

Inhaltsverzeichnis

	Vorwort	xi
1	Einleitung des menschlichen Autors	1
2	Automatisch erzeugte Inhalte: »Was zur Hölle ist das?«	5
3	Google, Facebook & Co.:	
	Was können und was planen die Big Five?	15
3.1	Google.....	17
3.2	Facebook, Apple, Microsoft & Amazon	27
4	»Automatische Textgenerierung für alle«:	
	Der templatebasierte Ansatz von Retresco	33
5	Roboterjournalismus:	
	Tod oder Zukunft einer Branche?	39
5.1	Wo ist die (künstliche) Intelligenz im automatisierten Journalismus?	40
5.2	Medienhistorische Vorläufer seit 1963.....	53
5.3	Was sagen die JournalistInnen, was sagen die LeserInnen?.....	57
5.4	Die Trends: Personalisiertes Storytelling, Dialogjournalismus, Sensorstorsys ...	62
6	Chatbots und Social Bots:	
	Wie »gute« und »böse« Bots die Kommunikation steuern	67
6.1	Chatbots, intelligente persönliche Assistenten und automatisierte Entscheidungen	68
6.2	Social Bots, Meinungsroboter und automatisierte Accounts.....	83
7	Automatisierte Kommunikation und die neue Kommunikationswissenschaft	95
8	Content-Automatisierung, Bildungssystem und Arbeitsmarkt-Prognosen	105
9	Ausblick: Die Automatisierung von allem?	113

10	Rechtliche Aspekte automatisch erstellter Inhalte	125
11	Glossar	131
12	Tabellen- und Abbildungsverzeichnis	135

Vorwort

Von Uwe Hannig

Die Automatisierung der Erstellung und nutzerorientierten Zurverfügungstellung von Inhalten über unterschiedliche Kanäle ist ein Wunschtraum nicht nur aller Content-Marketing-Manager. Dazu bedarf es allerdings noch weiterer Fortschritte im Forschungsfeld der künstlichen Intelligenz. Sprachgesteuerte, internetbasierte persönliche Assistenten, die Daten von Webseiten abrufen, verarbeiten und die Ergebnisse in natürlicher Sprache ausgeben, sind hier erst der Anfang. Die vollautomatische Erstellung von nutzerindividualisierten Artikeln wird allerdings noch einige Zeit auf sich warten lassen.

Werkzeuge zur Erleichterung des Content Management sind aber bereits verfügbar. Diese erlauben beispielsweise die Automatisierung der Publikation von Inhalten auf mehreren Plattformen. Heute geschieht dies in den meisten Fällen noch durch fehleranfälliges Kopieren und Einfügen. Aktualisierungen müssen, da die Dokumente nicht miteinander verbunden sind, in der Folge manuell in mehreren Medien durchgeführt werden. Aufgrund des Einsatzes von Dokumenten ohne Metadaten ist die Wiederverwendung von Inhalten für unterschiedliche Zielgruppen zeitaufwändig. Lösungen zur Automatisierung der Verwaltung und Veröffentlichung von Inhalten optimieren diese Prozesse¹ und erleichtern darüber hinaus die Übersetzung in andere Sprachen.

Neben der automatisierten Distribution und Wiederverwendung von Content in unterschiedlichen Medien bzw. auf verschiedenen Plattformen ist man bezüglich der Messung der Nutzung von Inhalten und der Zuordnung von Content zu Buyer Personas bzw. Zielkunden im Rahmen des Account-based Marketing entlang der Buyer Journey schon gut unterwegs. Das noch zu lösende Problem ist die erstmalige Erstellung von qualitativ anspruchsvollen

1) Dies meint bislang häufig »Content Automation«: Eher die automatisierte Verwaltung von Content in unterschiedlichen Dateiformaten und auf unterschiedlichen Plattformen als dessen tatsächliche Erstkreation/Erstellung, siehe etwa <http://quark.com>.

Inhalten. Diesbezüglich gilt heute noch: *Kein Inhalt ohne Autor*. Inhalte automatisch *beurteilen* lässt sich mittlerweile jedoch schon bewerkstelligen. So haben Forscher der Universität von Chicago eine Applikation entwickelt, die maschinell Rezensionen verfassen kann, die als ähnlich nützlich wie die von Menschen erstellten eingestuft wurden.²

Mittels KI fertigten Forscher der Rutgers-Universität Gemälde an, die es sogar weltweit in große Ausstellungen schafften und in einer Versuchsanordnung von Ausstellungsbesuchern besser bewertet wurden als Bilder von Menschenhand.³ Die Frage ist deshalb nicht mehr, *ob* mit Hilfe künstlicher Intelligenz in Zukunft nutzbare Inhalte erstellt werden, sondern *wie lange* es noch bis dahin dauert.

Dieses Buch erlaubt Ihnen einen überblickartigen Einstieg in die faszinierende neue Welt der Chatbots und Schreibroboter und deren Möglichkeiten. Kritisch hinterfragt werden aber auch die Auswirkungen der automatischen Content-Erstellung auf unser Bildungssystem und die Medienwissenschaft.

Viel Spaß bei der Lektüre wünscht Ihnen Ihr

Uwe Hannig

*Prof. Dr. Uwe Hannig ist wissenschaftlicher Leiter des
Instituts für Sales und Marketing Automation.*

-
- 2) Yuanshun Yao, Bimal Viswanath, Jenna Cryan u.a. (2017): Automated Crowdturfing Attacks and Defenses in Online Review Systems.
<https://people.cs.uchicago.edu/~ravenben/publications/pdf/crowdturf-ccs17.pdf>
 - 3) <https://sites.google.com/site/digihumanlab/home> sowie https://medium.com/@ahmed_elgammal/generating-art-by-learning-about-styles-and-deviating-from-style-norms-8037a13ae027

1 Einleitung des menschlichen Autors

»Why write your own book when an algorithm can do it for you?«¹

Dieses Buch wurde komplett *von einem Menschen* geschrieben. Idee zum Buch, Recherchen am Telefon, per E-Mail, vor Ort, in Büchern, Journals und im Web, Lektüre, quellenkritische Bewertung, Schreiben und Korrektorat – das alles hat ein Mensch durchgeführt. Die verwendeten technischen Hilfsmittel waren im Wesentlichen das Internet, E-Mail und MS Word. Dieser Vorgang könnte schon sehr bald der Vergangenheit angehören: Davon handelt dieses Buch. Die Idee, es zu schreiben, kam mir Ende 2017, als mich ein befreundeter Gastronom und Hotelier auf den Umstand hinwies, dass er den beschreibenden Text zu seinem Unternehmen auf Google Maps selbst nicht ändern kann. Versuche, Google per Kontaktformular auf Fehler aufmerksam zu machen, wurden vom Suchmaschinen Giganten beschwiegen. Ich zeigte den inkriminierten Text mehreren IT-JournalistInnen und WissenschaftlerInnen. Alle meinten, es sei ein computergenerierter Text, der sich aus Geodaten von Google Maps und Extrakten von Kundenbewertungen speist. Google verweigerte in der Folge Angaben über die Herkunft dieser und Tausender anderer Texte (»Hoteldetails«) auf Google Maps. – Ist das das transparente und freie Internet, das wir haben wollten?

-
- 1) [MED] Zu dieser Frage ein Interview mit Philip M. Parker: <https://readwrite.com/2013/01/15/why-write-your-own-book-when-an-algorithm-can-do-it-for-you>. Parker, im Hauptberuf Universitätslehrer, hält ein Patent auf die automatische Generierung von Büchern auf der Basis von Content aus dem Web und hat mit seiner Software nach Eigenangaben bereits mehr als eine Million Bücher zu Nischenthemen erstellt und rund 50.000 in seinem Verlag ICON Group auch publiziert, siehe [ECO] <https://www.icongrouponline.com> (am 30.08.2018 waren es exakt 49.389 Titel, vor allem automatisch generierte Vorschauen) sowie [MED] https://www.buffingtonpost.com/entry/philip-parker-book-s_n_2648820.html?section=india. Dabei greift Parker – ebenfalls Eigenangaben zufolge – auf ca. 60 Computer und ein Team von sechs Entwicklern zurück (Stand 2008), siehe [MED] <https://www.nytimes.com/2008/04/14/business/media/14link.html>. Parkers Long-Tail-Vorgehen der besonderen Art ist wissenschaftsethisch und urheberrechtlich umstritten. Zudem kann die genaue Anzahl »seiner« Bücher nicht verifiziert werden. (Long Tail: Eroberung von Nischenmärkten in den Netzmedien.)

Mir wurde schnell klar, dass hier etwas nicht stimmt: Google, das Unternehmen, das wie kein anderes unseren virtuellen Wirklichkeitszugang bestimmt, lässt die Herkunft von Beschreibungstexten über Tausende Unternehmen weltweit im Verborgenen. Im Ernstfall kann so ein Text über Erfolg oder Pleite eines Unternehmens und damit von Menschen entscheiden. Eine erste Recherche zeigte: Die automatische Erzeugung von Inhalten, die der Menschenhand entzogen sind, ist ein Thema, zu dem es noch kaum Literatur gab.

Ich versuche mit diesem Buch, die Frage zu beantworten, ob es sich bei derzeitigen Formen des *Automated Content* bloß um einen gegenwärtigen, womöglich schnell vergänglichen Hype oder um Vorboten der nächsten Stufe der digitalen Revolution handelt.

Zwei Hinweise zur Lektüre

Webquellen

Ich bin mir der *Vergänglichkeit von Webquellen* bewusst. Dennoch zitiere ich sie häufig. Kann eine zitierte Webquelle nicht mehr online vorgefunden werden, so existiert von ihr sehr häufig ein Snapshot, der in der Vergangenheit gemacht wurde. Bitte gehen Sie dazu auf die Webseite <https://archive.org> und kopieren Sie die hier im Buch zitierte URL in die Wayback Machine. Wählen Sie dann eine archivierte Version der Webseite unter einer Datumsangabe aus, die in der Nähe des Jahres 2018 ist. – Interessanterweise wissen immer noch viele Menschen über diese einfache Methode der Auffindbarkeit alter Webseiten nicht Bescheid.

Nicht-wissenschaftliche Quellen

Ich bin mir auch des Problems des *Zitierens nicht-wissenschaftlicher Quellen* bewusst. Mir ist dazu eine Lösung eingefallen, die ich bislang noch nirgendwo gesehen habe: Ich kennzeichne nahezu alle Quellen vor der Literaturangabe in der Fußnote mit ihrer Zugehörigkeit zu einem sozialen System (im Sinne Niklas Luhmanns): Genuin wissenschaftliche und populärwissenschaftliche Quellen erhalten den Verweis [SCI] für *science*. Massenmediale und journalistische Quellen, Social-Media- und Blogquellen erhalten den Verweis [MED] für *media*. Quellen aus der Wirtschaft wie Selbstdarstellungen von Unternehmen auf Webseiten sowie PR-Meldungen erhalten den Verweis [ECO] für *economics*. – Mittlerweile werden ja nicht nur in studentischen Qualifikationsschriften alle Quellen als gleichrangig betrachtet und

auch so verwendet. Mein Versuch, hier etwas Ordnung zu schaffen, kann freilich weiter ausdifferenziert werden, in juristische Quellen, politische Quellen usw. Ich freue mich auf kreative Nachahmer!

Mein Dank gilt folgenden Personen, mit denen ich die Inhalte dieses Buchs diskutieren durfte: Prof. DDr. Peter A. Bruck, Research Studios Austria Forschungsgesellschaft; Ao. Prof. Mag. Dr. Markus Haslinger, TU Wien; Dr. Mihai Lupu, Research Studios Austria Forschungsgesellschaft sowie Prof. Dr. Josef Mitterer, Universität Klagenfurt. Für das Vorwort zum Buch danke ich Prof. Dr. Uwe Hannig, für das juristische Schlusskapitel Mag. Dr. Albrecht Haller. Ich danke weiter meinen GesprächspartnerInnen Alexander Siebert von Retresco (Berlin), Frank Feulner von AX Semantics (Stuttgart) und Mag. Katharina Schell von der APA – Austria Presse Agentur (Wien) für die Zeit, die sie sich für meine Besuche bzw. mein wiederholtes Nachfragen genommen haben sowie meinen zahlreichen E-Mail-InterviewpartnerInnen weltweit. Mein Dank gilt schließlich auch Dr. Michael Barabas vom dpunkt-Verlag und Florian Rötzer von Telepolis. Christian Reifert und Susanne Rudi danke ich für das kompetente Lektorat und die professionelle Betreuung.

*Stefan Weber
Dresden und Salzburg,
Januar bis September 2018*