

# ÖFIT-Trendschau

Öffentliche Informationstechnologie in der digitalisierten Gesellschaft

Trendthema 33:

## **Denkende Maschinen**

Stand: Dezember 2015

**Herausgeber:**

Jens Fromm und Mike Weber  
Kompetenzzentrum Öffentliche IT  
Fraunhofer-Institut FOKUS  
Kaiserin-Augusta-Allee 31, D-10589 Berlin  
Telefon: +49 30 3463 - 7173  
Telefax: + 49 30 3463 - 99 - 7173  
info@oeffentliche-it.de  
www.oeffentliche-it.de  
www.fokus.fraunhofer.de

**Autorinnen und Autoren der Gesamtausgabe:**

Mike Weber, Stephan Gauch, Faruch Amini, Tristan Kaiser, Jens Tiemann,  
Carsten Schmoll, Lutz Henckel, Gabriele Goldacker, Petra Hoepner, Nadja Menz,  
Maximilian Schmidt, Michael Stemmer, Florian Weigand, Christian Welzel, Jonas Pattberg,  
Michael Rothe, Oliver Schmidt, Nicole Opiela, Florian Friederici, Jens Fromm

**Gestaltung:**

Reiko Kammer

ISBN: 978-3-9816025-2-4

Dezember 2015

*Bibliographische Angabe:*

Weber, Mike, Jens Tiemann, Christian Welzel, Lutz Henckel, Gabriele Goldacker,  
Petra Hoepner, Nadja Menz, Maximilian Schmidt, Michael Stemmer, Florian Weigand,  
Jonas Pattberg, Nicole Opiela, Florian Friederici, Jens Fromm 2015: Denkende Maschinen.  
In: Jens Fromm und Mike Weber, Hg., 2015: ÖFIT-Trendschau: Öffentliche Informationstech-  
nologie in der digitalisierten Gesellschaft. Berlin: Kompetenzzentrum Öffentliche IT.  
<http://www.oeffentliche-it.de/trendschau>.

Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 3.0 Deutschland  
Lizenz (CC BY 3.0 DE) <http://creativecommons.org/licenses/by/3.0/de/legalcode>. Bedingung  
für die Nutzung des Werkes ist die Angabe der Namen der Autoren und Herausgeber.

## Denkende Maschinen

Künstliche Intelligenz polarisiert wie kaum ein anderes Thema der Digitalisierung. Seit Jahrzehnten sollen »schon bald« nur schwer durchschaubare Maschinen den Menschen in seiner bislang exklusiven Domäne des problemlösenden Denkens übertrumpfen. Kinofilme nutzen diesen Angriff auf das menschliche Selbstverständnis ebenso für spektakuläre Darstellungen wie Feuilletons für philosophische und kulturkritische Betrachtungen. Seit einiger Zeit gewinnt das Thema durch konkrete Anwendungen neuen Schwung. Mit welchen Mechanismen und Auswirkungen werden wir in Zukunft tatsächlich konfrontiert? Verstricken wir uns in einem unkontrollierbaren Geflecht von Informationen mit unabsehbaren Folgen? »Siri, Google Now, Cortana – was meint ihr?«

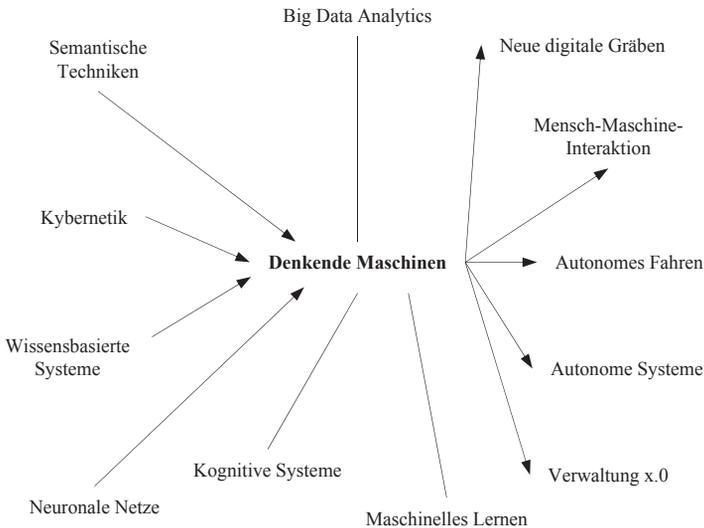


Abb. 1: Netzwerkartige Verortung des Themenfeldes: Vorläufer, Begleitphänomene und Folgen

Die Vision von denkenden Maschinen und künstlicher Intelligenz (KI) ist genauso alt wie die Computertechnik selbst. Immer wieder wurden schnelle, große Entwicklungssprünge versprochen. Bei allen beeindruckenden und technologisch wichtigen Einzelerfolgen – erinnert sei an Schachspiele und Quizshows – bleibt der Durchbruch in der Breite bisher aus. Derzeit sind wieder erstaunliche Fähigkeiten von Maschinen zu beobachten – etwa bei persönlichen Assistenten (s. Trendthema 26: Mensch-Maschine-Interaktion). Dabei können die Systeme auf bislang ungeahnte Datenmengen zur Kalibrierung und zum maschinellen Lernen zurückgreifen. Stehen wir nun kurz vor einem fundamentalen Durchbruch im Bereich KI?

Die Schwierigkeiten bei der Bewertung fangen schon bei der unklaren Begriffsdefinition für Intelligenz an. Wenig erstaunlich also, wenn Branchenkenner kognitive Systeme durch den »Wow-Faktor« von klassischen unterscheiden. Ein solcher Überraschungseffekt lässt sich nur schwer fassen. Menschen neigen dazu, Ereignissen Intentionen zuzuschreiben. Darunter haben nicht nur Haustiere zu leiden, sondern es erklärt auch den bemerkenswerten Erfolg so einfacher Programme wie etwa den legendären Chatbot ELIZA aus dem Jahr 1966, dem mitunter auf seine allgemeinen Fragen sehr persönliche Informationen offenbart wurden. Zielführend ist die Beschreibung von kognitiven Systemen anhand von vier Eigenschaften, die – mehr oder weniger stark ausgeprägt – allesamt vorliegen müssen: Kontextualität (Verständnis für die reale und virtuelle Umgebung in einer Situation), Adaptivität (Lernfähigkeit im veränderlichen Umfeld), Interaktivität (komfortable Interaktion mit Nutzern) und Iterativität (Dialogfähigkeit zum Präzisieren von Zielen und Problemen).

Die KI-Forschung verfolgt zwei grundsätzliche, sich idealiter befruchtende Linien: Einerseits ist das Ziel, mehr über die ungeklärten Vorgänge des menschlichen Denkens zu erfahren. Für diesen Ansatz steht etwa das aktuell stark diskutierte Human Brain Project, das durch die Nachbildung neuronaler Prozesse des Gehirns explizit auf vermarktungsfähige Forschungsergebnisse zielt. Andererseits geht es um clevere Algorithmen, mit denen Computer Alltagsprobleme lösen können. Auch ohne Verständnis der Mechanismen menschlichen Denkens werden so nützliche Lösungen entwickelt. Mustererkennung mittels sogenannter neuronaler Netze oder die Erfassung von Semantik aufgrund statistischer Verfahren sind dafür Beispiele.

Den Unterschied zu klassischen, stärker determinierten Systemen machen Verfahren zum maschinellen Lernen aus: Damit können technische Systeme selbsttätig grundsätzliche Eigenschaften oder Muster erkennen und in der Folge auch auf bis dahin

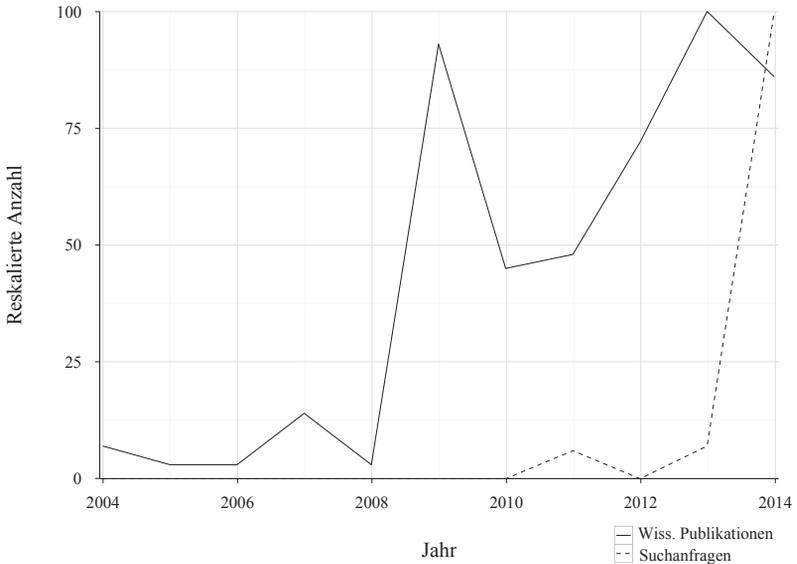


Abb. 2: Wissenschaftliche, wirtschaftliche und gesellschaftliche Themenkonjunkturen (zum Stichwort »Cognitive Computing«)

unbekannte Eingangsdaten und in einem breiten Einsatzbereich anwenden. Hierfür wird »Wissen« in Form des internen Zustands des Systems aufgebaut.

Zur vollständigen Beherrschbarkeit eines Systems gehört neben der Kenntnis der Eingangswerte auch der seines internen Zustandes. Anhand der Daten allein lassen sich die Ergebnisse nicht mehr reproduzieren. Kontrolle ist also nur bei kontinuierlichem Zugang zum System möglich. Die Folgen sind umfangreich: Das aus den Daten gewonnene Wissen verbleibt verdeckt im System. Auch bei festgelegten Rahmenbedingungen führt ein definierter Anfangszustand nicht zwangsläufig zu einem reproduzierbaren Endzustand. Vielmehr muss das Systemverhalten durch komplexe Teststellungen auf Basis typischer Einsatzsituationen bewertet und gegebenenfalls unter Zuhilfenahme statistischer Verfahren evaluiert werden. Die Möglichkeit zur vollständigen Systemabbildung bleibt dabei für Dokumentation und forensische Analysen wünschenswert, reicht für die Analyse eines sich ständig selbst anpassenden Systems aber nicht aus.

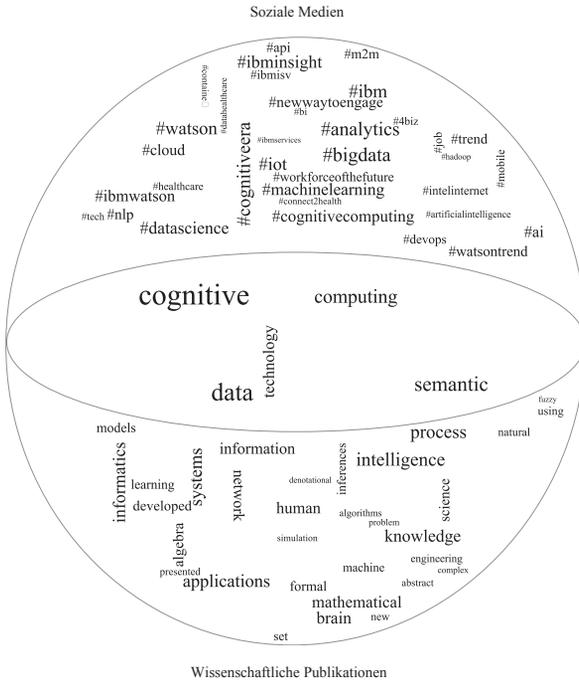


Abb. 3: Häufigkeit von Hashtags bei Twitter und von Stichworten in wissenschaftlichen Publikationen (zum Stichwort »Cognitive Computing«)

Klassische Ansätze zur Beherrschbarkeit stoßen bei kognitiven Systemen also schnell an ihre Grenzen. Umso wichtiger ist es, die grundsätzlichen Eigenschaften der verwendeten Algorithmen zu kennen. Ein Systembestandteil sollte eine Funktion für die Beschreibung der Herleitung seiner Ergebnisse sein. Zwar obliegen solche Beschreibungen wiederum undurchsichtigen Algorithmen, sie sorgen jedoch für eine bessere Nachvollziehbarkeit der sonst als von einer Blackbox generiert empfundenen Endergebnisse. Gewissenhaft umgesetzt können solche Selbsterklärungsfunktionen nicht nur zu einem besseren Verständnis der Systeme, sondern auch zu einer sinnvollen Anpassung der Zielfunktionen führen. Im Idealfall werden so Zielvorstellungen und Zwischenschritte zu deren Erreichung transparenter und leichter gestaltbar.

Mögliche Folgen für das Gemeinwesen hängen stark von der Art der kognitiven Systeme, der Betrachtungsebene und dem Anwendungsfeld ab. Die Übernahme trivialer Alltagsaufgaben dürfte bei hinreichender Systemintegrität unkritisch sein. Auch dürften einzelne komplexe Teilaufgaben wie autonomes Fahren (s. Trendthema 25) so gut zu bewältigen sein, dass sie insgesamt zu einer höheren (Verkehrs-)Sicherheit führen. Je individueller das Ergebnis, je folgenschwerer das Anwendungsfeld und je autonomer die kognitiven Systeme, desto problematischer werden Fragen nach Verantwortung und Haftung für die möglicherweise schwerwiegenden Folgen. Ein zweiter Punkt betrifft die Überwindung bestehender und Schaffung neuer digitaler Gräben (s. Trendthema 8). Einerseits erlauben kognitive Systeme einen Grad an Nutzerfreundlichkeit, der eine Teilhabe breiterer Bevölkerungsgruppen verspricht; andererseits können neue digitale Gräben entstehen, wenn die Teilhabe faktisch vertiefte technische und inhaltliche Kenntnisse erfordert. Entscheiden Computersysteme über menschliche Schicksale mit – etwa in der Bildung oder bei der Arbeitsplatzsuche – kann eine unangemessene Nutzung auch tragische Folgen haben.

Noch lösen Computer spezielle, klar umrissene Aufgaben mittels beträchtlicher Datenmengen und weiterhin wachsender Rechenleistung. Bereits damit erreichen sie beeindruckende Erfolge. Wenn sich dereinst die Verheißungen quasi-menschlich denkender Maschinen erfüllen, stellen sich wesentliche Fragen neu.

Möglichkeiten	Wagnisse
<ul style="list-style-type: none"> <li>– Realisierbarkeit komplexer technischer Systeme</li> <li>– Anpassungsfähigkeit von Maschinen an Mensch und Umwelt</li> <li>– Schnittstellenerweiterung</li> <li>– Verstärker menschlicher Kreativität</li> <li>– Kontinuierliche, sich selbst verbessernde Entscheidungsunterstützung</li> <li>– Transparente Entscheidungsabläufe</li> <li>– Werkzeug für Wissensarbeit</li> <li>– Übernahme von Routine-Wissensarbeiten</li> </ul>	<ul style="list-style-type: none"> <li>– Beherrschung der Komplexität und Vernetzung natürlicher oder technischer Systeme</li> <li>– Entwicklung von Konzepten und Werkzeugen zum Test komplexer Algorithmen</li> <li>– Interpretation theoriefreier Analysen</li> <li>– Entstehung neuer digitaler Gräben</li> <li>– Verantwortung und Verantwortungszuschreibung bei negativen Folgen</li> <li>– Umbrüche des Arbeitsmarktes</li> </ul>

*Handlungsraum a: Forschungsförderung zu den Themen Testen und Selbsterklären*  
Um die Zuverlässigkeit und Korrektheit kognitiver Systeme zu testen, reichen Systemabbildungen nicht mehr aus. Ergänzend bedarf es komplexer Teststellungen und statistischer Analysen. Zugleich müssen kognitive Systeme ihre eigenen Ergebnisse nachvollziehbar begründen können. Hierzu bedarf es umfassender interdisziplinärer Forschung aus einer Gemeinwohlperspektive.

*Handlungsraum b: Neue Formen der Markt- und Sicherheitsregulierung*  
Die Eigenschaften kognitiver Systeme machen neue Regulierungsregime notwendig. Bestehende Fragen nach Rechten an Daten werden durch Fragen nach Rechten am gewonnenen Wissen verschärft. Gleiches gilt für die Gewährleistung der Sicherheit bei Entwicklung und Betrieb derartiger Systeme.

*Handlungsfeld c: Anwendungen im öffentlichen Sektor*  
Kognitive Systeme bieten sowohl in der öffentlichen Leistungserstellung (s. Trendthema 29: Verwaltung x.0) als auch zur Fundierung politischer Entscheidungen beträchtliche Potenziale. Diese zu nutzen und dabei neue Formen der Transparenz zu entwickeln, kann Effizienz und Effektivität zum Vorteil der Bürgerinnen und Bürger steigern.

