Digitalisierung

Claus Harringer*

„Good Bot, Bad Bot“?

Zur Problematik von Bot-Ontologien

<https://doi.org/10.1515/iwp-2018-0040>

Zusammenfassung: Zu „Bots“ wird viel geforscht und geschrieben, meist jedoch in sehr kleinteiliger Form, die die größeren Kontexte außer Acht lässt, in denen sich diese selbsttätigen Programme bewegen. Auch Typologien finden sich kaum – wenn doch, dann stammen sie eher von Bot-Enthusiast_innen. Eine umfassende Ontologie zum Thema Bots leistet auch dieser Beitrag nicht, er versucht aber, durch Kritik an bestehenden Bot-Ontologien ein Kriterium zu entwickeln, mit dem sie bewertet werden können. Dieses Kriterium ist die Empfehlung, alle Bot-Ontologien zu ignorieren, die Bots nicht mit Big Data oder KI-Systemen in Verbindung bringen.

Deskriptoren: Bot, Automatisierung, Big Data, Künstliche Intelligenz, Ontologie, Informationssystem, Datenbank, Thesaurus, Taxonomien

Abstract: Although there is currently much research and publishing on „Bots“, most of it is rather unsystematic and generally disregards the larger contexts these automated software agents operate in. Typologies are also rare and if they exist, they are mostly created by enthusiasts. The claim of this article however is neither to provide a comprehensive ontology of „bots“, nor to lay out the groundwork for such an undertaking, but to develop a criterion for the evaluation of existing ontologies. In a nutshell this criterion consists of the recommendation to ignore all bot-ontologies that neglect to consider the connection between bots, artificial intelligence and big data.

Deskriptoren: Bot, Automatisierung, Big Data, Künstliche Intelligenz, Ontologie, Informationssystem, Datenbank, Thesaurus, Taxonomien

Résumé: On effectue de nombreuses recherches et on écrit énormément sur les « bots », mais surtout sous une forme très fragmentée qui ignore les contextes plus larges dans

lesquels ces programmes autonomes fonctionnent. Même les typologies sont difficiles à trouver – et si on en rencontre, il s’agit surtout du fruit de passionnés de bots. Cet article ne fournit pas non plus une ontologie complète des bots, mais il essaie de développer un critère d’évaluation en étudiant de façon critique les ontologies de bot existantes. Ce critère consiste dans la recommandation d’ignorer les ontologies de bots qui n’associent pas les bots à Big Data ou à des systèmes d’intelligence artificielle.

Descriptors: Bot, Automatization, Big data, Artificial intelligence, Ontologies, Information systems, Databases, Thesauri, Taxonomies

Bereits eine oberflächliche Auseinandersetzung mit dem Thema „Bots“ führt vor Augen, dass diese in mannigfaltigen Erscheinungsformen auftreten. Umso bemerkenswerter ist dabei, dass es kaum Versuche gibt, sie umfassend zu kategorisieren. Wenn diese dennoch unternommen werden, sind sie – weil oftmals von deklarierten Bot-Enthusiast_innen stammend – leider meist von Distanz- und Kritiklosigkeit gekennzeichnet. Auch die wissenschaftliche Literatur zum Thema trägt wenig zur systematischen Erfassung des Phänomen „Bots“ bei – Typologien finden sich darin kaum, dafür aber zahlreiche Studien zu *bestimmten* Auswirkungen *spezieller* Bots, ohne, dass umfassender auf Beziehungen zwischen ihnen und vor allem zu ihrer Umwelt reflektiert wird. Eine umfassende Bot-Systematik kann und will auch dieser Beitrag nicht leisten – im Folgenden geht es aber um den Versuch, vermittels einer kritischen Prüfung existierender Bot-Systematiken am Ende zu einem Kriterium zu gelangen, mithilfe dessen sich deren Brauchbarkeit einschätzen lässt.

Der Titel des Textes enthält – abseits einer Anspielung auf polizeiliche Verhörtaktiken („Good Cop – Bad Cop“), deren Bedeutung sich noch erschließen wird – Material für drei Fragen:

1. Was ist „Ontologie“?
2. Was sind „Bots“?
3. Was sind die Schwierigkeiten bei der Erstellung von Bot Ontologien?

*Kontaktperson: Claus Harringer, Johannes Kepler Universität Linz, Institut für Philosophie und Wissenschaftstheorie, Altenberger Straße 69, 4040 Linz, Österreich, E-Mail: claus.harringer@jku.at

1 Was ist „Ontologie“?

1.1 Ontologien in der Philosophie

Zunächst ist zu klären, was mit „Ontologie“ gemeint ist. Der Begriff kommt aus der Philosophie und bedeutet – je nach Übersetzung – „Lehre vom Sein“ oder „Lehre vom Werden“. Es gibt diverse Formulierungen für Definitionen und jedes philosophische Lexikon wird einen jeweils anderen Aspekt betonen.¹ Ich habe mich für eine recht prägnante Bestimmung entschieden, die sich auf der Homepage der Philosophischen Fakultät der Universität München findet:

„Die Ontologie beschäftigt sich mit allem, was es gibt, denn sie fragt erstens, was es heißt, daß es etwas gibt, und zweitens, welche Kategorien von Objekten existieren und in welchem Verhältnis sie zueinanderstehen.“²

Als Arbeitsgrundlage ist das ausreichend – die wirklichen Probleme ergeben sich dann, wenn es darum geht, eine Ontologie zu erarbeiten, die mit möglichst wenigen Kategorien auskommt, zugleich aber so viel wie möglich erfasst. Dieses Problem stellte sich bereits Aristoteles und es ist auch heute noch Gegenstand philosophischer Anstrengungen.

Der Verweis auf die philosophische Herkunft von Ontologie dient an dieser Stelle vorrangig dazu, den Bedeutungswandel des Begriffs zu verdeutlichen, den er in der Übertragung auf andere Bereiche und Disziplinen durchgemacht hat: Den Kern ontologischer Überlegungen bildete in der Philosophie traditionell die Reflexion auf die Bedeutung der verwendeten Kategorien. Dagegen wird Ontologie in Dokumentation, Bibliothekswesen und Informationswissenschaften, sowie Informatik und Computerwissenschaften – bei allen Unterschieden – vorwiegend von ihrer Ordnungsfunktion her begriffen. Dies zeigt sich nicht zuletzt in den Verwendungen von Ontologien bei Datenbanken und anderen Informationssystemen.

1.2 Ontologien in Dokumentation, Bibliothekswesen und Informationswissenschaften

Ontologische Fragen beschäftigen Dokumentation, Bibliothekswesen und Informationswissenschaften in Grund-

zügen bereits seit dem 19. Jahrhundert: Ein bekanntes Beispiel ist die *Dewey-Dezimalklassifikation* zur inhaltlichen Erfassung von Bibliotheksbeständen. Dieser lange Verwendungszeitraum hat jedoch nicht unbedingt dazu geführt, dass eine eindeutige Auffassung über Charakter und Anwendungsbereiche von Ontologien besteht – zugespitzt formuliert ist der ontologische Status von Ontologien hier unterbestimmt. Im Folgenden seien deshalb nur einige Aspekte angedeutet:

Mit Frederico Fonseca ließe sich von einer „doppelten Rolle“ von Ontologien in der informationswissenschaftlichen Forschung sprechen, die klarer differenziert werden müsste: Ontologien *von* Informationssystemen („ontologies of information systems“) sind dazu da, die Werkzeuge zur Bildung von Begriffsmodellen zu bewerten. Sie beschäftigen sich damit auf einer grundlegenden konzeptionellen Ebene mit Informationssystemen im Allgemeinen. Ontologien *für* Informationssysteme („ontologies for information systems“) sind solche, bei denen es um die *Nutzung bestimmter Informationssysteme* geht, die in computerbasierter Form vorliegen und mit denen deren Komponenten erzeugt werden können (Fonseca, 2007). Gerade durch das Wegfallen technischer Limitationen in Darstellung und Übertragung von Dokumenten stehen Ontologien bei der Erstellung von Datenbanken vor neuen Herausforderungen, die auf konzeptioneller Ebene zu berücksichtigen sind.³

Eine weitere Schwierigkeit in der Auseinandersetzung mit Ontologien in den Informationswissenschaften besteht darin, dass oftmals Ontologien verwendet, diese aber nicht als solche bezeichnet werden. Richtig unübersichtlich wird es schließlich, wenn man versuchen will, sie von anderen relevanten Termini zu unterscheiden, wie „Thesaurus“ und „Taxonomie“. Mit „Ontologie“ haben diese beiden Begriffe gemein, dass sich ihre Bedeutung durch Aneignung und Gebrauch in den Informationswissenschaften stark gewandelt hat. Zusammen mit den mit dem „Web 2.0“ aufgekommenen „*Folksonomies*“ (bei denen die Nutzer_innen die Verschlagwortung durchführen) lassen sie sich als vier Formen *kontrollierter Vokabulare* fassen.⁴

¹ Eine übersichtliche Darstellung der Begriffsgeschichte findet sich etwa in Mittelstraß (2016), S. 35f.

² <http://www.philosophie.uni-muenchen.de/fakultaet/schwerpunkte/ontologie/index.html> [12.8.2018].

³ So ist etwa die Größe einer Datenbank gemessen an der Anzahl von Einträgen nur mehr bedingt aussagekräftig. Mit dem Aufkommen von Datenbanksystemen, die nicht mehr nur indixierten, sondern auch Abstracts enthielten, war eine große Datenbank nicht mehr unbedingt besser als eine kleinere, die Abstracts enthielt (Jaksó, 2007).

⁴ Ein kontrolliertes Vokabular ist eine festgelegte Liste von Wörtern, die zum Indexieren verwendet wird. Dadurch wird einerseits eine gewisse Einheitlichkeit der Kategorisierung erreicht, was andererseits aber auch das Erlangen relevanter Informationen bei der Suchabfrage erleichtert (du Preez, 2015).

Thesauri enthalten Indexbegriffe, die den Inhalt eines Dokuments beschreiben und funktionell die Rolle eines Instruments zu Kontrolle und Standardisierung der Terminologie erfüllen. Die Begriffe stehen dabei in einer hierarchischen Beziehung, die logisch von breiteren zu engeren Ausdrücken fortschreitet. Sie stehen aber auch in Verbindung mit anderen Begriffen auf der gleichen hierarchischen Ebene. Darin dürfte der Hauptunterschied zu *Taxonomien* liegen: Bei diesen herrscht zwischen den Begriffen eine *strikt* hierarchische Ordnung. Diese besteht bei Ontologien in der Form nicht, da sie zusätzlich bestimmte Eigenschaften von Begriffen mitberücksichtigen (du Preez, 2015). Ein weiterer wichtiger Unterschied besteht darin, dass Taxonomien und Ontologien mittels Informationstechnologien erstellt werden, während Thesauri auch manuell konstruiert werden können. Ontologien und Taxonomien sind damit prädestiniert für automatisierte Verfahren von Indexierung einerseits und Informationsabfrage andererseits. Während der Fokus in Dokumentation, Bibliothekswesen und Informationswissenschaften auf der möglichst umfassenden Beschreibung von Entitäten aus verschiedenen Bereichen und unterschiedlichen Blickwinkeln liegt, interessieren sich Computerwissenschaften und Informatik im Umgang mit Dokumenten vorrangig für logische Schlussfolgerungen.

1.3 Ontologien in Computerwissenschaften und Informatik

In den Computerwissenschaften ist der Begriff „Ontologie“ seit den 1960er Jahren im Bereich der Wissensrepräsentation (*knowledge representation*), einem Subfeld der „Künstlichen Intelligenz“ (KI) in Verwendung. Er bezeichnet darin eine allgemeine Struktur von Begriffen, die durch ein logisches Vokabular repräsentiert werden. Seit den 1990er Jahren bezeichnet er auch eine Sammlung von Technologien, die unter dem Etikett „semantisches Netz“ zusammengefasst wurden. Die neuen Möglichkeiten, die sich damit boten, sorgten für eine Art Renaissance der KI-Forschung, da das semantische Netz maschinelle Schlussfolgerungen (*automated reasoning*) erleichtert.

Mittlerweile gibt es in den Computerwissenschaften eine große Zahl an Publikationen zu Ontologien: Meist werden sie dabei als Hilfsmittel verstanden, das bei der Modellierung – sowohl von Datenbanken, als auch der Wissensrepräsentation – zur Anwendung kommt. Der bekannteste Definitionsversuch von Ontologien in der Informatik stammt vom Computerwissenschaftler und Unternehmer Thomas Gruber: Er bezeichnete Ontologie in einem Artikel Anfang der 1990er Jahre als eine „explizite

Spezifikation einer gemeinsamen Konzeptualisierung“ (Gruber, 2013). Diese Bestimmung scheint in ihrer Allgemeinheit wenig aussagekräftig – konkreter formuliert geht es bei Ontologien um eine standardisierte Terminologie, sowie Ableitungsregeln zwischen dort definierten Begriffen (Hesse, 2002). Zusammengefasst erfüllen Ontologien in der Informatik einerseits die Funktion einer logikbasierten Theorie über die sich Bereiche verstehen und auf Modelle reduzieren lassen (worin die philosophische Grundbedeutung noch mitschwingt), als auch andererseits die eines Hilfsmittels zur Erstellung von systemspezifischen Vokabularen und der Erzeugung automatischer Schlussfolgerungen (Almeida, 2013). Das Stichwort Automatisierung leitet damit direkt über zum Thema „Bots“.

2 Was sind „Bots“?

2.1 Arbeitsdefinition „Bots“

Begriffsgeschichtlich lässt sich der Ausdruck „Bots“ über „Roboter“ und das alttschechische „robota“ (Zwangsarbeit) bis zum altkirchenslawischen Ausdruck für Sklave („*rab*“) zurückführen. Diese etymologische Volte der Übertragung menschlicher Zwangsverhältnisse auf Maschinen (und in weiterer Folge Software), ist zwar anekdotisch aufschlussreich, erhellt aber kaum die wesentlichen Eigenschaften von „Bots“. Ganz allgemein lassen sie sich als Software charakterisieren, die entweder automatisch oder unter minimaler menschlicher Einwirkung Befehle ausführt, auf Nachrichten antworten oder Routineaufgaben durchführt. Explizit zu betonen ist dabei, dass ihr Einsatzgebiet das Internet ist und auch automatisierte Accounts in Social Networks darunter fallen.

2.2 Anzahl „Bots“

Es lässt sich nicht sagen, wie viele Bots in absoluten Zahlen im Internet unterwegs sind – nicht zuletzt weil dies mit der Schwierigkeit verbunden ist, die Frage „Bot, or not bot?“ klar zu beantworten. Die Internet-Sicherheitsfirma Incapsula hat zwischen 2012 und 2016 einen jährlichen „*Bot-Traffic Report*“ herausgegeben, bei dem sie versucht hat, das Bot-Aufkommen statistisch zu untersuchen. Der bislang letzte Report aus dem Jahr 2016 analysierte 16.7 Milliarden Besuche auf 100.000 zufällig ausgewählten Domains innerhalb des Incapsula-Netzwerkes, um unter anderem zu erfahren, wie viel Datenverkehr von Bots generiert wird. Im Längsverlauf zeigt sich, dass *nur in einem*

der fünf erfassten Jahre Bots nicht mehr als die Hälfte des Internettraffics auf diesen Seiten bestritten.⁵

Inwieweit diese Verhältniszahlen generalisierbar sind, ist strittig: Zum einen, weil das „Incapsula-Network“ eine Vorauswahl darstellt, die nicht zum Zwecke einer wissenschaftlichen Untersuchung getroffen wurde. Zweitens wäre zu prüfen, welche Arten von Bots auf diesen Seiten in Erscheinung traten – schließlich entscheiden Ausrichtung und Beschaffenheit der Website-Auswahl auch darüber, welche Bots auf ihnen agieren.

2.3 Genese „Bots“

Als wichtiger Vorläufer heutiger Bots lässt sich das Programm ELIZA benennen, das sich mit Möglichkeiten der Mensch-Computer Kommunikation über natürliche Sprache beschäftigte und Mitte der 1960er-Jahre vom Informatiker Joseph Weizenbaum am MIT entwickelt wurde. Der Wortschatz des Programms stammt aus einem Thesaurus – einem strukturierten Wörterbuch mit systematisch geordneten Ausdrücken. ELIZA simuliert eine psychotherapeutische Sitzung, in der die Nutzer_innen auf Fragen antworten. Als Grund für dieses Setting gab Weizenbaum an, dass eine psychotherapeutische Sitzung eine der wenigen Gesprächssituationen ist, in der ein Teil vorgeben kann, nichts von der Welt zu wissen und unbekümmertes Nachfragen das Gegenüber nicht irritiert. Dieses Nachfragen entfaltet dabei insofern therapeutische Wirkung, als sie bei Nutzer_innen des Programms durchaus den Eindruck entstehen lässt, sich *tatsächlich* mit einem Menschen zu unterhalten.

2.4 Arten von Bots

In Form einer ungeordneten Liste ließen sich als wichtigste Bot-Arten folgende anführen:

- *Webcrawler* (*Search Engine Spiders*, bzw. *Robots*) sind Programme, die sich durchs Netz bewegen und Webseiten indexieren, um sie für Suchanfragen verfügbar zu machen. Sie werden aber auch dafür eingesetzt, Informationen zu sammeln
- *Monitoring Bots* überwachen und kontrollieren die Funktionsfähigkeit von Software
- *Impersonators* verschaffen sich Passwörter, indem sie sich tarnen
- *Messenger Bots* sind Kommunikationsmodule auf Messengerplattformen
- *Scrapers* holen sich E-Mail-Adressen von Webseiten
- *Spammers* schwemmen Websites mit Reklame, um den Datenverkehr umzulenken und Klicks zu generieren
- *Hackers* attackieren Websites und deponieren Schadsoftware auf ihnen
- *Botnets* wiederum sind eine ganze Armee von Computern die mit automatisierter Schadsoftware infiziert wurden und zentral gesteuert für dezentralisierte Attacken genutzt werden
- *Chat Bots/Chatterbots* stellen aus Sicht der Nutzer_innen die vorherrschende Botart dar: Sie sind dazu programmiert, möglichst perfekt menschliche Kommunikation zu imitieren. Einer ihrer Hauptanwendungsbereiche liegt im Kund_innenservice
- *Social Bots* werden meist für den Versuch eingesetzt, bestimmte Zielgruppen durch automatisch generierte Inhalte zu beeinflussen. Die Bots bedienen sich dabei oft falscher oder gehackter Accounts in sozialen Medien, die sich für wenig Geld erwerben lassen. Für Social Bots, bei denen es um politische Einflussnahme geht, hat sich auch die Bezeichnung *Political Bots* etabliert
- *Transaction Bots* wiederum interagieren mit externen Systemen, um Daten von einer Plattform auf eine andere zu übertragen – sie forcieren damit die Automatisierung von Geschäftsprozessen. Wenn es sich dabei um Hochfrequenzhandel („high frequency trading“) handelt, heißen sie *Trading bots*. Da sich Bots in der Geschäftswelt und der Kommunikation am Arbeitsplatz generell durchsetzen, lässt sich auch von *Work Bots* sprechen.
- *Editing Bots* erleichtern die Arbeit bei Wikipedia
- *Assistant Bots* sind schließlich die „All-in One“-Variante unter den Bots, die sich von der Wiege bis zur Bahre um alles kümmern

3 Bot-Ontologien und deren Beschränkungen

Nun wäre es durchaus möglich, diese Bots danach zu ordnen, welche Funktionen sie erfüllen. Es ließe sich aber auch ganz einfach konstatieren, dass sie alle das aufweisen, was der Philosoph Ludwig Wittgenstein als „Familienähnlichkeit“ bezeichnet hat: Damit ist gemeint, dass Mitglieder einer taxonomischen Gruppe ein Bündel ähnlicher Eigenschaften teilen, ohne dass diese allen Mitgliedern zukommen müssen. Wenn man sich mit den wenigen

⁵ <https://www.incapsula.com/blog/bot-traffic-report-2016.html> [12.8.2018].

vorhandenen Bot-Typologien beschäftigt, fällt allerdings auf, dass vor allem eine Eigenschaft als Kriterium herangezogen wird, an das man aus wissenschaftlichen Überlegungen heraus nicht unbedingt denken würde.

3.1 „Good“ Bots vs. „Bad“ Bots

Auf Plattformen und Websites von Bot-Enthusiast_innen findet sich häufig das duale Schema einer Unterteilung in „Good Bots“ und „Bad Bots“.⁶ Dabei werden etwa *Chatbots*, *Crawlers*, *Transactional Bots* und *Informational Bots* den „Guten“ zugeschlagen, während *Hackers*, *Spammers*, *Scrapers*, und *Impersonators* zu den „Bösen“ gezählt werden. Dieser Zugang findet sich auch beim zuvor angesprochenen „Bot-Traffic Report“.

Aus diesem ist – unter den bereits zuvor formulierten Vorbehalten gegenüber der Aufstellung – auch ersichtlich, dass die „bösen“ Bots, die „guten“ quantitativ übertreffen. Die IT-Sicherheitsfirma „distilnetworks“ gibt sogar jährlich einen eigenen „Bad Bot Report“ heraus.⁷ Wenig überraschend findet sich das Schema „Good Bot – Bad Bot“ oftmals auf Seiten von IT-Sicherheitsfirmen. Es lässt sich damit nicht zuletzt als Reklamestrategie begreifen.

Das augenscheinlichste – aber nicht einmal größte – Problem bei dieser Zuordnung besteht darin, dass Bots eine moralische Qualität zugesprochen wird. Der wirklich kritische Punkt ist aber der, dass bei der Unterteilung in „Good Bots“ und „Bad Bots“ *kategoriale Zuordnungen* verwendet werden, wo *relationales Denken* gefragt wäre – anstelle Bots qua Gruppenzugehörigkeit einer der beiden Sorten zuzuschlagen, wären deren jeweiligen Kontexte in Betracht zu ziehen. An zwei kleinen Beispielen sei demonstriert, dass die absolute Einteilung in „gut“ und „böse“ nicht immer so einfach ist.

Bad „Spam Bot“?

Spam Bots werden für gewöhnlich als schädliche Angelegenheit betrachtet – eine Ausnahme dürfte der „Impostor Buster“ darstellen. Der Bot agierte ausschließlich auf Twitter und wurde vom Journalisten Yair Rozenberg initiiert. Der Anlass für den Impostor Buster war, dass Rozenberg während seiner Berichterstattung über die Trump Prä-

sidentschaftskampagne 2016 mit Massen antisemitischer Kommentare konfrontiert wurde. Daraufhin beschäftigte er sich näher mit den Online-Strategien der sogenannten Alt-Right-Bewegung. Ein wichtiger Teil davon sind „Impersonator Trolls“. Das sind Menschen, die falsche Identitäten annehmen, um bestimmte Gruppen zu diffamieren. Derart „getarnt“, mischen sie sich dann in Diskussionen ein und versuchen, diese zu eskalieren. Der Spam Bot „Impostor Buster“ hatte nun die Aufgabe, diese Trolle unschädlich zu machen. Dazu wurde er mit einer sorgfältig zusammengestellten Datenbank verbunden, in die derartige Troll-Accounts eingepflegt worden waren. Sobald sich eines dieser Profile in eine Debatte einmischte, trat der Bot in Aktion. Das klingt nach einer Sache, die durchaus im Interesse des Plattformbetreibers – in diesem Fall Twitter – sein müsste. Allerdings wurde der Impostor Buster Ende 2017 von Twitter wegen „Harassment“ gesperrt, weil die von ihm betroffenen Trolle erfolgreich eine Spam-Beschwerde eingelegt hatten.⁸

Good „Chat Bot“?

Am 23. März 2016 wurde auf Twitter der Account eines Chatbot mit Namen „Tay“ eröffnet und bereits tags darauf wieder entfernt. Das Ganze hatte hoffnungsvoll mit dem obligaten „hello world!“ begonnen – von diesem ohnehin bescheidenen Niveau aus ging es dann aber recht schnell bergab. Am 24. März 2016 twitterte Tay Sachen wie „I fucking hate feminists and they should all die and burn in hell“, sowie „Hitler was right I hate the jews“. Microsoft entschuldigte sich, der Schaden fürs Image (und auch die Schadenfreude der Konkurrenz) war aber beträchtlich. Die Erklärung für diese Entwicklung hat einerseits mit der gewählten Plattform, viel mehr aber noch mit unterschiedlichen Internetkulturen zu tun. Vorbild für Tay war ein Chat-Bot namens *XiaoIce*, den Microsoft für den chinesischen Markt entwickelt hatte. Tay selbst wurde auch nicht primär für Twitter entworfen, sondern dort erst relativ spät ausgesetzt. Er stellte ein Experiment dar, wie ein Chat-Bot die natürliche Sprache einer bestimmten Gruppe erlernt: Im konkreten Fall handelte es sich um die Alterskohorte der 18- bis 24-Jährigen, speziell in den USA. Das Problem dabei war, dass das chinesische und das US-amerikanische Onlinesystem recht unterschiedlich sind: Chinesische Onlineplattformen sind stark reglementiert, US-amerikanische nicht. Vor allem gibt es in China keine Message

⁶ Als Beispiel für eine durchwegs affirmative und unkritische Typologie kann die bienenfließige Zusammenstellung auf der Seite „Botnerds“ gelten (<http://botnerds.com/types-of-bots/> [12.8.2018]).

⁷ <https://resources.distilnetworks.com/white-paper-reports/2018-bad-bot-report> [12.8.2018].

⁸ <https://www.nytimes.com/2017/12/27/opinion/digital-nazi-hunter-trump.html> [12.8.2018].

Boards, auf denen sich Menschen verabreden, gezielt auf Bots loszugehen und zu versuchen, sie zu stören und unbrauchbar zu machen. Für Microsoft war Tay einerseits ein Fehlschlag, zugleich aber Lernprozess: In Folge beschäftigten sich die Entwicklungsteams damit, Mechanismen zu entwickeln, die verhindern sollten, dass der Bot mit Aussagen gefüttert wird, die er nicht lernen soll. Das Resultat ist Tays Nachfolger „Zo“, den Microsoft im März 2017 veröffentlichte und der nach wie vor aktiv ist.

Gegen diese Argumentation ließe sich natürlich einwenden, dass diese Fälle lediglich Ausnahmen darstellen, die eine pragmatische Einteilung in „gute“ und „böse“ Bots nicht grundsätzlich in Frage stellen. Der Hinweis auf weitere Beispiele (die sich ohne weiteres anführen ließen), wäre als Entgegnung allerdings wenig hilfreich, weil es dann lediglich darum ginge, die existierende Bot-Typologie differenzierter zu gestalten. Dadurch würde aber der eigentliche Punkt verfehlt: Eine ethische Bewertung von Bots ist auf dieser Ebene nicht sinnvoll, weil sie den erweiterten Kontext außer Acht lässt, in dem sie sich bewegen.

3.2 Bots & Künstliche Intelligenz

Mit dem Beispiel des Chatbot „Tay“ ist implizit bereits ein wichtiges Kriterium zur Klassifikation von Chatbots angesprochen: Die Lernfähigkeit. Am unteren Ende der Skala finden sich dabei *Script Bots* (wie ELIZA): Diese führen nur das aus, was in ihnen als algorithmische Anweisungen codiert ist. *Smart Bots* sind dagegen in der Lage, von sich aus auf umfassende Datensammlungen zuzugreifen und auf Situationen flexibel zu reagieren. An der Spitze der Intelligenzpyramide findet sich dann das, was als *Intelligent Agents* bezeichnet wird. Damit sind wir schließlich im Bereich der KI angelangt und es sei hier nur kurz angedeutet, was unter „Machine Learning“, „Deep Learning“ oder „Selbstverstärkendem Lernen“ verstanden werden kann: 2015 hat Google im Rahmen seines „DeepMind“ Projekts ein Experiment veröffentlicht, das recht anschaulich zeigt, wohin die Reise geht. Losgelassen wurde „DeepMind“ auf ein altes Atari-Videospiel, bei dem es darum geht, mit einem Ball Blöcke abzuschießen. DeepMind wurde allerdings nur die Vorgabe gegeben, den Punktestand zu vergrößern. Ansonsten wurde vorab nichts „erklärt“: Weder, was ein Spiel ist, noch wie es funktioniert, noch – und das ist am wichtigsten – was all die Daten, mit denen es konfrontiert wird, überhaupt repräsentieren (Formen, Farben, Bewegung etc.). Dies alles musste das System durch stumpfes Ausprobieren und Abgleichen der Veränderungen selbst herausfinden. Erst stellte es sich noch „un-

geschickt“ an – nach zwei Stunden „spielte“ es bereits professionell und nach vier Stunden hatte es die optimale „Strategie“ entwickelt, um die meisten Punkte zu erzielen.

3.3 Bots, Big Data und Künstliche Intelligenz

Bislang war von verschiedenen Bots und deren Kategorisierung die Rede und es scheint nun so, als wären die KI-getriebenen Bots lediglich eine besonders leistungsfähige Form, die eventuell andere, weniger mächtige ablösen könnte. An dieser Stelle ist aber zu betonen, dass Bots und KI-Systeme mittlerweile einen Gesamtzusammenhang bilden, bei dem eine Binnendifferenzierung wenig sinnvoll ist, weil sie den Blick auf die wesentlichen Tendenzen im digitalen Bereich verstellt. Komplettiert wird diese Troika durch „Big Data“, dessen wesentliche Eigenschaften *Geschwindigkeit* (Erzeugung von Daten in Echtzeit) und *Vollständigkeit* ($n=alle$) sind.⁹ Vielleicht verleitet auch bereits die bloße Identifikation eines Programms als „Bot“ zum nominalistischen Fehlschluss, sich nur auf solche Entitäten zu konzentrieren, die als „Bots“ benannt werden. An dieser Stelle sei deshalb vorgeschlagen, eher von „Automatisierungsagenten“ zu sprechen, da Automatisierung letztlich die große Klammer bildet, welche die diversen Tendenzen eint. Dies sei abschließend noch kursorisch am Feld der Wissenschaften exemplifiziert.

3.4 Die Automatisierung der Wissenschaften

Zwar gibt es Beispiele dafür, explizite „Science Bots“ in der wissenschaftlichen Arbeit einzusetzen,¹⁰ diese stellen jedoch nur einen kleinen Ausschnitt dessen dar, was sich hinsichtlich der Automatisierung der Wissenschaften anführen ließe. Ulrich Herb hat am Beispiel des „Betriebssystems“ Elsevier eindrucksvoll die ubiquitäre Durchdringung des wissenschaftlichen Alltags durch Altmetrics beschrieben, die faktisch unentrinnbar jeden Schritt von den ersten Forschungsskizzen über Durchführung und Publikation bis hin zu den Reaktionen auf die resultierenden Artikel (Zitierungen, Medienerwähnungen) alles misst und kommerziell zu vernutzen sucht (Herb, 2018). Derart generierte Daten gehen wiederum eine Mésalliance mit der

⁹ Diese beiden Charakteristika identifizierten Kitchin/McArdle (2016) in einem Vergleich von 26 Datensätzen als die relevantesten, noch vor Umfang („Volume“) und Vielfalt („Variety“).

¹⁰ Tobias Kuhn schlug 2015 den Einsatz von „Science Bots“ vor. Ihre Aufgaben wäre die automatische Verarbeitung von Daten, sowie deren Publikation (Kuhn, 2015).

herrschenden „audit culture“ ein, die damit auch Evaluationen, Mittelvergabe und die scheinbar unvermeidlichen „Rankings“ auf automatisierte Basis stellt. Über das Ausmaß, in dem automatisierte Verfahren auch in der Wissenschaftskommunikation zum Einsatz kommen (etwa beim Verfassen von Abstracts), kann nur spekuliert werden, im Wissenschaftsjournalismus sind Bots jedenfalls bereits etabliert (Tatalovic, 2018).

Auch für (wissenschaftliche) Bibliotheken wird seit Jahren der verstärkte Einsatz von Chatbots debattiert (Dee Ann; 2012) und nicht zuletzt die oben erwähnten Möglichkeiten automatisierter Indexierung von Fachpublikationen betreffen unmittelbar den Charakter wissenschaftlicher Recherche.

Auch die Lehre an Hochschulen ist betroffen: Automatisierte Plagiatsprüfungen sind nur unter der Prämisse personeller Unterversorgung als Entlastung der Lehrenden zu begreifen – als kompensatorische Maßnahme helfen sie dabei, den geforderten „Output“ (gemessen in Abschlusszahlen) zu gewährleisten.

Im Kern getroffen wird Wissenschaft letztlich dann dadurch, dass die „Automatisierungsagenten“ jene Entwicklung forcieren, die Chris Anderson (2008) im Aufkommen von Big Data angelegt sah und begrüßte: Das „Ende der Theorie“, in dem wissenschaftliche Modell- und Theoriebildung durch die algorithmische Potenz des „n=alle“ ersetzt wird.

Damit lässt sich die Entstehung einer Billigwissenschaft prognostizieren, die – gleichsam selbst zum Bot geworden – automatisch oder unter minimaler menschlicher Einwirkung Befehle ausführt. Wissenschaftliche Erkenntnis wird damit ersetzt durch deren Simulation im Sinne effekteischender Reproduktion bereits etablierten Wissens.

4 Quines Messer

Als Fazit dieses kleinen Streifzugs möchte ich eine Faustregel im Umgang mit Bot-Ontologien vorschlagen, die ich als „Quines Messer“¹¹ bezeichne:

„Ignoriere alle Definitionen und Klassifikationen von Bots, die diese nicht in einen größeren technologischen und gesellschaftlichen Zusammenhang stellen.“

¹¹ Der analytische Philosoph Willard Van Orman Quine etablierte 1969 das ontologische Prinzip der „natural kinds“. Damit ist gemeint, dass einer Gruppe von Objekten bestimmte Eigenschaften zukommen – diese sind jedoch nicht Produkt der Klassifikation, sondern kommen den Objekten natürlicherweise zu (Quine, 1969).

Das heißt konkret, dass sich der Erkenntnisgewinn in der Auseinandersetzung mit Bots daran bemisst, inwieweit sie sich nicht darauf beschränkt, kleinteilige Sortierarbeit vorzunehmen und Bots wie Sammelbilder zu behandeln. Es kommt einerseits darauf an, größere technologische Rahmenbedingungen mitzudenken und sich zu fragen, welche Bedeutung Bots in Verbindung mit „Big Data“ und „Künstlicher Intelligenz“ haben. Eine separate Behandlung, die diese Einbettung ignoriert, produziert fragmentierte Spezialdiskurse und blockiert zugleich das Verstehen von Zusammenhängen. Dies ist besonders folgenreich, wenn es um die sozialen und ethischen Implikationen technologischer Umbrüche geht. Letztendlich sollte eine ethisch motivierte Auseinandersetzung mit Bots als „Automatisierungsagenten“ darauf hinauslaufen, sie danach zu klassifizieren, inwieweit sie wiederum die Tendenz aufweisen, ihrerseits Menschen zu klassifizieren – sei es nach Wahl-, Konsum-, Forschungs- oder Sozialverhalten. Dementsprechend sind alle involvierten Wissenschaften aufgefordert, die hier angedeuteten Kontexte in ihrer Komplexität ernst zu nehmen und so gesellschaftlichen Debatten eine fundierte Basis zu geben.

Literatur

- Almeida, M. B. (2013); Revisiting ontologies: A necessary clarification; *J Am Soc Inf Sci Tec*; 64; 1682-1693; DOI: 10.1002/asi.22861.
- Anderson, C. (2008); The end of theory: The data deluge makes the scientific method obsolete; *Wired*, 23 June 2008; via http://www.wired.com/science/discoveries/magazine/16-07/pb_theory [30.8.2018].
- Dee Ann, Allison (2012); Chatbots in the Library: Is it Time?; *Library Hi Tech* 30.1; 95-107; DOI: 10.1108/07378831211213238.
- Fonseca, F. (2007); The double role of ontologies in information science research; *J. Am. Soc. Inf. Sci.*; 58; 786-793; DOI: 10.1002/asi.20565.
- du Preez, M. (2015); Taxonomies, folksonomies, ontologies: what are they and how do they support information retrieval?; *The Indexer*; 33; 29-37.
- Gruber, T. A. (1993); Translation Approach to Portable Ontology Specifications; *Knowledge Acquisition*; 5(2); 199-220.
- Herb, U. (2018); Zwangsehen und Bastarde; *Information – Wissenschaft & Praxis*; 69(2-3); 81-88; DOI: 10.1515/iwp-2018-0021.
- Hesse, W. (2002); *Ontologie(n)*; Informatik-Spektrum; 25; 477; DOI: 10.1007/s002870200265.
- Jacsó, P. (2007); How big is a database versus how is a database big; *Online Information Review*; Vol. 31; Issue: 4; 533-536; DOI: 10.1108/14684520710780467.
- Kitchin, Rob / McArdle, Gavin (2016); What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets; *Big Data & Society*; 3; DOI: 10.1177/2053951716631130.
- Kuhn, T. (2015); Science Bots: A Model for the Future of Scientific Computation?; *WWW 2015 Companion Proceedings*; <http://dx.doi.org/10.1145/2740908.2742014> [30.8.2018].

- Mittelstraß, J. (Hg.) (2016); Enzyklopädie Philosophie und Wissenschaftstheorie; Band 6: O – Ra; Stuttgart: J. B. Metzler.
- Quine, Willard V. (1969); Natural kinds; In: Jaegwon Kim & Ernest Sosa (eds.); Ontological Relativity and Other Essays; Columbia University Press; 114–38.
- Tatalovic, M. (2018); AI writing bots are about to revolutionise science journalism: we must shape how this is done; JCOM; 17 (01); E.; DOI: 10.22323/2.17010501.

Anmerkung: Dieser Text basiert auf einem Vortrag, der am 27. April 2018 auf dem Symposium „BOTS or NOT BOTS? Politik, Demokratie & Wahlen im Zeitalter ihrer digitalen Produzierbarkeit“ an der Johannes Kepler Universität (JKU) Linz gehalten wurde. Veranstaltet wurde das Symposium vom Kulturinstitut an der JKU zusammen mit dem Institut für Philosophie und Wissenschaftstheorie mit Unterstützung der Österreichischen Gesellschaft für Dokumentation und Information (ÖGDI).

Claus Harringer

Johannes Kepler Universität Linz
 Institut für Philosophie und Wissenschaftstheorie
 Altenberger Straße 69
 4040 Linz
 Österreich
claus.harringer@jku.at

Claus Harringer B. A., M. A. ist Lektor am Institut für Philosophie und Wissenschaftstheorie der Johannes Kepler Universität Linz.