

Pre-Print Version

Forthcoming in:

DFJV (Hrsg.) (2015): Journalistische Genres.
Konstanz: UVK.

Citation:

Dörr, Konstantin (2015): Algorithmic Journalism.
Einordnung und Konsequenzen. In: DFJV (Hrsg.)
(2015): Journalistische Genres. Konstanz: UVK
[im Erscheinen].

0 Algorithmic Journalism: Einordnung und Konsequenzen

„Wir halten es weder für wahrscheinlich noch für wünschenswert, dass journalistische Formate von Algorithmen bestimmt werden“¹, so der Deutsche Journalisten-Verband im Jahr 2014. Die Realität sieht anders aus. Algorithmen übernehmen bereits journalistische Funktionen, indem sie editieren, aggregieren, distribuieren und automatisiert publizieren.² Software ist „our interface to the world, to others, to our memory and our imagination [...]“³ und der Computer längst fester Bestandteil der Nachrichtenproduktion im 21. Jahrhundert. Diese Veränderungen zeigen sich im Journalismus bereits bei der Nachrichtensuche (z. B. Google), der Nachrichtempfehlung (z. B. reddit) oder der Nachrichtenaggregation (z. B. Rivva).⁴

Doch viele Journalisten, Redakteure, Betriebsräte und Verlage sind nervös. Die „Roboter-Journalisten“ kommen. „Could a Computer Write This Story?“, fragt der US-amerikanische Nachrichtensender CNN und die englische Tageszeitung The Guardian stimmt mit „The Robot Journalist: Heralding an Apocalypse for the News Industry?“ gar endzeitliche Töne an.

Zwar war und ist Automatisierung immer schon Antriebsfeder des Journalismus,⁵ doch der Kampf um Profitabilität, Marktanteile, Leser und publizistisches Renommee ist in Zeiten des Medienwandels erbitterter denn je. Aktuell drängen Firmen wie Automated Insights, Narrative Science, Arria, Yseop, Linguastat oder, in Deutschland, aexea und text-on mit ihren Produkten auf das journalistische Hoheitsgebiet der Inhaltserstellung und weisen auf ein neues Rollenbild hin: „For the first time journalists are migrating from a direct to an

¹ Deutscher Journalisten-Verband (DJV) im August 2014 auf E-Mail-Anfrage des Autors zum journalistischen Leitbild und der Integration algorithmischer Inhaltserstellung und „Robo-Journalisten“ in das journalistische Selbstverständnis.

² Vgl. Pavlik (2000; 2013); Gillespie (2014); Broussard (2014); Diakopoulos (2014).

³ Vgl. Manovich (2013), S. 2.

⁴ Vgl. Gillespie (2014); Wallace und Dörr (2014).

⁵ Vgl. Stöber (2013); Örnebring (2010).

indirect role.“⁶ Teile der Literatur sprechen gar von einer „silent marginalization of professional journalism within public communication.“⁷ Nachrichtenagenturen wie der Sport-Informationen-Dienst (SID), Thomson-Reuters oder die Associated Press (AP), aber auch traditionsreiche Zeitungsverlage wie die New York Times Company oder Hubert Burda Media experimentieren bereits mit Inhalten, die von Algorithmen „geschrieben“ werden.⁸

Damit nimmt die Relevanz algorithmischer Textgenerierung für die strategische und wirtschaftliche Ausrichtung vieler Medienorganisationen und den täglichen Medienkonsum von Tag zu Tag zu.⁹ Mit dieser Entwicklung hat der Journalismus ein beispielloses Maß an Automatisierung der Medienproduktion erreicht. Dabei ist die algorithmische Textgenerierung ein Produkt, das an der Schnittstelle zwischen Journalismus und der wachsenden Bedeutung der statistischen Datenanalyse aus „Big Data“ entstanden ist.¹⁰

Versuche, die Trends der Digitalisierung, Computerisierung und Automatisierung sowie den Einfluss von Algorithmen auf den Journalismus in einheitliche Konzepte zu fassen, liefern stets unterschiedliche Ergebnisse. Genres wie Data Journalism¹¹, Computational Journalism¹² oder Computer-Assisted Reporting¹³ – um nur einige zu nennen – fasst die norwegische Kommunikations- und Medienwissenschaftlerin Astrid Gynnild deshalb allgemein unter dem Phänomen der „computational exploration in journalism“¹⁴ zusammen. Dabei nimmt der Einfluss von Algorithmen auf Prozesse der gesellschaftlichen Kommunikation stetig zu. Bernhard Rieder, Associate Professor für

⁶ Vgl. Napoli (2014), S. 350.

⁷ Für einen Überblick vgl. Donsbach (2014), S. 661; Gynnild (2014).

⁸ Hier u.a.: Ein Pilotprojekt des SID zur automatischen Erstellung von Terminankündigungen von Sportveranstaltungen; AP und Reuters automatisieren Meldungen zu Geschäftsberichten und Geschäftszahlen von US-Unternehmen; der Quakebot & Homicide Report der LA Times; Live-Analysen von NFL Spielen durch den „4th Down Bot“ der New York Times; der Feinstaub-Monitor der Berliner Morgenpost oder das Finanzportal Finanzen100.de von Hubert Burda Media (in Partnerschaft mit Focus Online).

⁹ Vgl. Napoli (2014).

¹⁰ Vgl. Carston (2014), S. 2; Mayer-Schönberger und Cukier (2013).

¹¹ Vgl. Gray, Boungegru und Chambers (2012).

¹² Vgl. Hamilton und Turner (2009).

¹³ Vgl. Mayo und Leshner (2000).

¹⁴ Vgl. Gynnild (2014).

Neue Medien an der Universität von Amsterdam, verweist dabei auf die Entwicklung von Algorithmen hin zu Ordnungseinheiten, die Datensammlung und -analyse sowie Entscheidungsprozesse selbstständig durchführen können.¹⁵

Dennoch unterscheidet sich Algorithmic Journalism – speziell in der Funktionsweise, der Rolle der Algorithmen und im Ergebnis – elementar von anderen journalistischen Genres. Während Algorithmen dort hauptsächlich als Analyse-, Editier- oder Recherchewerkzeuge, zum Beispiel zur Visualisierung oder zur Darstellung von Zusammenhängen aus Big Data, verwendet werden, ermöglicht Algorithmic Journalism – nach produkt- und ergebnisabhängigen Trainingsrunden der Software – eine unabhängige Produktion, Publikation und Distribution von Texten (fast) ohne menschlichen Einfluss.

Dabei prägen Begriffe wie „robotic“ oder „robo journalism“, „automated journalism“, „computer/machine-written journalism“ oder „algorithmic journalism“¹⁶ den wissenschaftlichen und medialen Diskurs. Für die Beschreibung und Auseinandersetzung der automatisierten Textgenerierung im Journalismus wird der Begriff „Algorithmic Journalism“¹⁷ verwendet, da – wie gezeigt wird – Anwendungen auf Basis algorithmischer Selektion „often contain prefixes such as ‘algorithmic’ or simply ‘algo’ [...]“¹⁸

Zunächst gilt es dabei, das Genre in den allgemeinen Rahmen der algorithmischen Selektion im Internet einzuordnen, von der algorithmischen Textgenerierung ohne journalistischen Mehrwert abzugrenzen und zu definieren. Die Darstellung und die Diskussion der technischen Funktionsweise münden in eine Diskussion über Chancen und Gefahren des Genres für den Journalismus und in Anreize für den weiteren Diskurs.

¹⁵ Vgl. Rieder (2014).

¹⁶ Vgl. van Dalen (2012); Anderson (2012; 2013); Carlson (2014); Latzer et al. (2014); Napoli (2014); Clerwall (2014); Coddington (2014); van der Kaa und Kraemer (2014); Young und Hermida (2014).

¹⁷ Vgl. Anderson (2012).

¹⁸ Vgl. Latzer et al. (2014), S. 4.

Algorithmische Selektion, Textgenerierung & Algorithmic Journalism

Nach Latzer et al. ist „Algorithmische Selektion im Internet“ der technisch-funktionale Kern einer Fülle erfolgreicher Softwareanwendungen, z. B. von Google, Facebook, Amazon oder Netflix. Das Erfolgsgeheimnis dieser Anwendungen ist nichts anderes als die Lösung von Problemen entlang der Komponenten Eingabe, Verarbeitung und Ergebnis (Input – Throughput – Output). Dabei findet eine Auswahl von Elementen aus einer Gesamtheit statt.¹⁹ Die Herausforderung – so Latzer et al. – bestehe dabei in einer Auswahl und Filterung der jeweiligen Informationen sowie in einer Strukturierung, Ordnung und Sortierung je nach Anwendungsgebiet und Zweck.

Diese beiden Aspekte – die Selektion (Auswahl) und die Relevanzzuweisung (Gewichtung) – spiegeln sich in der Funktionsweise algorithmischer Selektion und deren Anwendungen wider. Gibt man bei der Google-Suche den Begriff „Süddeutsche Zeitung“ ein, erwartet man Informationen über oder aus dieser Zeitung und nicht über deren Konkurrenz aus Frankfurt.

Diese grundlegende Funktionsweise lässt sich auch auf die Form der algorithmischen Textgenerierung übertragen. Während die Textgenerierung (Natural Language Generation – NLG) in der Computerlinguistik schon auf eine jahrelange Tradition verweisen kann²⁰, gilt auch hier das Prinzip der Selektion von Daten, der Relevanzzuweisung bestimmter Eigenschaften in eine semantische Struktur und des Outputs in Form der Generierung eines Textes in natürlicher Sprache. Die Computerlinguistik spricht hier von den Stufen der Inhaltsfestlegung und Organisation (Input), der Textplanung, Aggregation, Lexikalisierung, Satzplanung und Oberflächenrealisierung (Throughput) und einem Text als Endprodukt (Output).²¹ Dabei ist Textgenerierung nach Reiter und Dale „automatically producing human (natural) language from a computational representation of information.“²²

Anbieter dieser Lösungen auf Basis von NLG, wie Automated Insights, Narrative Science oder aexea, sind mit ihren Produkten verstärkt im E-Commerce tätig und erschließen erst allmählich den Markt mit journalistischen

¹⁹ Vgl. Latzer et al. (2014), S. 4.

²⁰ Vgl. Reiter und Dale (2000); Carstensen et al. (2010).

²¹ Vgl. Carstensen et al. (2010), S. 437.

²² Vgl. Reiter und Dale (2000).

Angeboten. Mit wenigen Ausnahmen²³ sind diese Produkte entweder in einer Experimentierphase oder erst im Übergang zur Expansionsphase im Markt und hauptsächlich im Bereich der Finanz- oder Sportberichterstattung zu verorten.

Doch algorithmische Textgenerierung ist nicht auch zwangsweise immer Journalismus. So ist die automatisierte Erstellung von Geschäftsberichten, Marktanalysen, Patientenberichten im Krankenhaus, Produkttexten im E-Commerce oder Gefahrenvorhersagen (z. B. auf Ölplattformen) zwar Textgenerierung im Sinne der Computerlinguistik, Journalismus ist das aber nicht.

Hier wird eine institutionalistische Sichtweise²⁴ nach Kiefer²⁵ vertreten, die Journalismus als zentrale Funktion der Demokratie versteht. Auch die Tatsache, dass es zahlreiche verschiedene praktische und theoretische Zugänge und Definitionen zum Journalismus gibt²⁶, macht es notwendig, Algorithmic Journalism ausgehend von journalismusimmanenten Funktionen, Werten und Normen zu diskutieren.

Dieser institutionellen Konzeption folgend ist Algorithmic Journalism hier als (semi-) automatisierter Prozess der algorithmischen Textgenerierung zu verstehen. Zentral ist dabei die Selektion elektronischer Daten (Eingabe/Input), eine Gewichtung und Relevanzzuweisung verschiedener Daten-Eigenschaften in eine semantische Struktur (Verarbeitung/Throughput) und die Publikation des journalistischen Textes in natürlicher und grammatikalisch korrekter Sprache. Der Text ist auf einer Online- oder Offline-Plattform mit publizistischer Reichweite (Ergebnis/Output) sichtbar, wird innerhalb oder außerhalb einer Redaktion professionell nach bestimmten Leitlinien und Werten produziert, erfüllt die Kriterien der Aktualität, Publizität, Periodizität und Universalität und stellt damit Öffentlichkeit her.

Die technischen Anforderungen und die Funktionsweise des Genres lassen sich durch die Darstellung in einem E-V-E-Modell²⁷ veranschaulichen. Dabei

²³ Vgl. Fußnote Nr. 8.

²⁴ Vgl. hier auch Napoli (2014), der Algorithmen als Institutionen diskutiert.

²⁵ Vgl. Kiefer (2010), S. 163.

²⁶ Vgl. Neuberger (2002); Malik (2004); Meier (2011).

²⁷ Das Eingabe-Verarbeitung-Ergebnis-Modell ist ein Grundprinzip der Datenverarbeitung; in der englischen Literatur arbeitet man mit den Begriffen Input – Throughput – Output; Latzer et al. (2014); Lewis und Westlund (2014).

werden die Ebenen und Prozesse der Textgenerierung aus der Computerlinguistik und dessen Grenzen – speziell im Hinblick auf eine journalistische Textproduktion – berücksichtigt und anschließend diskutiert.

E-V-E-Modell Algorithmic Journalism

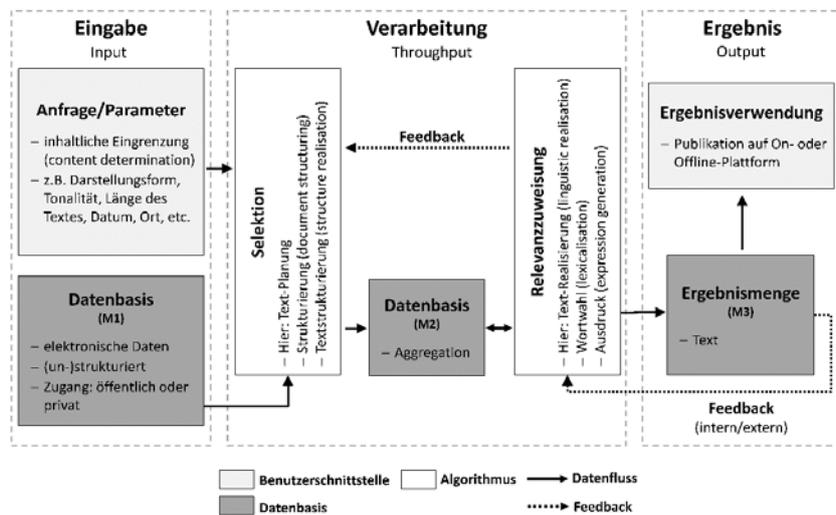


Abb. 1: Eingabe-Verarbeitung-Ergebnis-Modell Algorithmic Journalism in Anlehnung an Latzer et al. (2014), Lewis und Westlund (2014), Reiter und Dale (2000) und Carstensen et al. (2010), eigene Darstellung.

Ausgangspunkt für algorithmischen Journalismus sind zwei Mengen: eine Datenbasis (M1), über die operiert wird, und eine Ergebnismenge (M3), die anfangs leer ist. Für die Textgenerierung und die Verarbeitbarkeit ist es notwendig, dass die Daten in elektronischer Form strukturiert oder unstrukturiert (textuell) vorliegen. Der Zugang kann über öffentliche Programmierschnittstellen (APIs) oder über private Datenbanken erfolgen, die extern zugekauft werden. Die Daten sind dabei die journalistische Währung, der Rohstoff für jede

Form der Textgenerierung mit journalistischer Zielsetzung, aber auch Anknüpfungspunkt für ethische Fragestellungen²⁸, wie die nach der Objektivität des Datensatzes.

Für die Textgenerierung ist auf bzw. vor dem Input-Level eine Vorstrukturierung und inhaltliche Eingrenzung notwendig.²⁹ Dabei geht es um Parameter wie Länge des Textes, journalistische Darstellungsform, Thema, Tonalität/Stil sowie Ort und Zeit der Publikation. Diese Eingrenzung ist notwendig, da ohne sie die Textgenerierung an ihre technischen Grenzen stößt. So ist die Inhaltserstellung in klar abgrenzbaren Domänen (Kontexten) wie z. B. Wetter, Verkehr, Finanzthemen oder Sport aktuell ebenso möglich wie die Produktion von Texten in unterschiedlichen Sprachen, Textlängen und Stilen, je nach Programmierung. Sobald allerdings der inhaltliche Rahmen zu breit wird (z. B. Darstellungsform Reportage), kämpft die Computerlinguistik mit Ambiguitätsproblemen, also Mehrdeutigkeiten von Wörtern oder Sätzen (Beispiel Ball: Lederkugel oder Tanzveranstaltung?).³⁰ Eine selbständige journalistische Einordnung oder Reflexion, z. B. zu politischen Themen, stellt den Algorithmus (noch) vor unüberwindbare Herausforderungen. Was nicht im Vorfeld inhaltlich als Regelwerk im Programmcode definiert wurde, kann auch nicht als Ergebnis abgebildet werden. Der Output steht also größtenteils bereits fest. Je länger und inhaltlich komplexer ein Text sein soll, desto schwieriger sind demnach auch Programmierung und Darstellung.

Im Throughput finden die Textplanung und die Textrealisierung statt. Im Auswahlprozess vergleicht dabei ein Selektionsalgorithmus die Elemente der Datenbasis (M1) iterativ mit den dafür festgehaltenen Bedingungen. Erfüllt ein Element die festgelegten Bedingungen, wird es in die Zwischenmenge (M2) aufgenommen. Erfüllt es sie nicht, wird es davon ausgeschlossen. Eine darüber hinausgehende Ordnung, Strukturierung und Sortierung ist damit jedoch noch nicht notwendigerweise verbunden. Diese erfolgt mit der Relevanzzuweisung der Elemente der Zwischenmenge (M2). Hier finden die Textrealisierung, die Aggregation und die Lexikalisierung statt. Daraus resultiert als Ergebnis eine – nach Relevanzkriterien sortierte – Ergebnismenge in Form eines Textes in grammatikalisch korrekter, natürlicher Sprache (M3). Die Relevanzzuweisung erfolgt dabei durch eine automationsgestützte, statistische

²⁸ Vgl. hier den Abschnitt *Chancen und Gefahren des Algorithmic Journalism – Objektivität des Datensatzes*.

²⁹ Vgl. Reiter und Dale (2000), S. 49.

³⁰ Vgl. Carstensen et al. (2010); Reiter und Dale (2000).

Bewertung der Elemente nach festgelegten – programmierten – Relevanzkriterien.³¹

Das Modell zeigt auch, dass algorithmischer Journalismus klar von Nachrichten- oder Content-Aggregatoren wie Google News oder Summly abzugrenzen ist, da bei diesen Anwendungen das Element der originären Inhaltserstellung fehlt und nur bereits bestehende Inhalte zusammengefasst werden. Jedoch ist die Textgenerierung ein höchst komplexer Prozess in der Computerlinguistik und man darf nicht davon ausgehen, dass Algorithmen selbstständig, kreativ und ohne menschlichen Einfluss lesbare Texte „schreiben“ können.³² Feedbackschleifen signalisieren den menschlichen Einfluss auf den Generierungsprozess, der so lange optimiert wird, bis das erwünschte Ergebnis in Form eines Textes sichtbar wird.

Dabei arbeiten z. B. für Firmen wie Narrative Science oder aexea sowohl Texter, Journalisten als auch Computerlinguisten im Produktionsprozess eng zusammen. Texter schreiben Satzbausteine und Beispieltex te je nach gewünschter Tonalität und gewünschtem Inhalt, die anschließend durch Computerlinguisten nach semantischen Logiken angepasst und in Programmcode übersetzt werden. Diese Anpassungen müssen für jedes gewünschte Produkt, für jede Textart durchgeführt werden. So entsteht im Lauf der Zeit ein Regelwerk unterschiedlicher Satzkonstruktionen und Formulierungen.

Diese veränderten Produktionsroutinen werfen auch die Frage auf, ob Journalisten in Zukunft vertiefte Informatikkenntnisse benötigen und ob deren Vermittlung in eine universitäre und praktische Journalistenausbildung integriert sein sollte. Das Beispiel des US-Journalisten Ken Schwenke – er programmierte den Quakebot für Erdbebenvorhersagen der Los Angeles Times – zeigt, dass IT- und Programmierkenntnisse stark nachgefragt sind. Schwenke wechselte für weitere Projekte zu der New York Times.

³¹ Vgl. hier den Abschnitt *Chancen und Gefahren des Algorithmic Journalism* – journalistische Werte im Programmcode.

³² Vgl. Carstensen (2010).

Chancen und Gefahren des Algorithmic Journalism

Diese Entwicklungen zeigen, dass algorithmischer Journalismus kein isoliertes Phänomen, sondern Teil einer immer weiter fortschreitenden Rationalisierung, Automatisierung und Simulierung der menschlichen Kommunikation in der Nachrichtenindustrie ist, der Medienorganisationen in ihrer strategischen Ausrichtung beeinflusst. Schon jetzt sind oft nicht mehr nur journalistische Qualität oder publizistische Ziele, sondern vor allem der wirtschaftliche Ertrag ausschlaggebend.³³

Dennoch sind Rationalisierungsargumente und journalistische Beißreflexe nur bedingt begründet, da die noch sehr faktenbasierte Textgenerierung an sprachliche und darstellerische Grenzen stößt und im Augenblick nur für spezielle Themenbereiche – beispielsweise aus der Finanz- oder der Sportwelt, etwa Börsen- oder Ergebnisticker – genutzt werden kann. Zudem ist die Textgenerierung stark von der Qualität und Struktur der Daten abhängig. Auf einer rein technischen Ebene zeigt sich aber, dass Algorithmic Journalism institutionelle Aufgaben des Journalismus durchaus erfüllen kann.³⁴

Grundsätzlich dient algorithmischer Journalismus als Ausgangsbasis für weitere journalistische Recherchen und inhaltliche Weiterentwicklung und reduziert so Transaktionskosten. Dabei kann über Special-Interest-Themen berichtet werden, die vielleicht nur für ein Nischenpublikum interessant sind. Aufgrund der Skalierbarkeit, der möglichen Sprachenvielfalt algorithmisch erstellter Texte und der kostengünstigen Produktion³⁵ wären sie damit wirtschaftlich rentabel.

Erste Rezeptionsstudien³⁶ zur Qualität, zur Glaubwürdigkeit und zur Wahrnehmung algorithmisch erstellter Texte zeigen, dass Unterschiede zwischen menschlich geschriebenen und automatisierten Texten nur schwer auszumachen sind. Van Dalen stellt in einer qualitativen Journalistenbefragung sogar fest, dass Journalisten „Robo-Journalismus“ auch als Chance sehen,

³³ Vgl. Stavelin (2013), S. 27; Kiefer (2001), S. 33.

³⁴ Eine Potentialanalyse der Technologie zeigt, dass journalistische Prinzipien und Kriterien der Aktualität, Periodizität, Universalität und Publizität erfüllt werden können.

³⁵ Vgl. www.ax-semantic.com. Hier erhält man für 250 Euro bis zu 100 Texte in einer Länge zwischen 250 und 350 Wörtern; aktuell sind Texte in 10 Sprachen möglich.

³⁶ Vgl. van der Kaa und Kraemer (2014); Clerwall (2014).

„[...] when routine tasks can be automated, journalists will have more time for in-depth reporting“³⁷.

Unter ethischen Gesichtspunkten ergeben sich für den Journalismus zentrale Fragen nach Herkunft und Objektivität der Daten, die für die Textgenerierung verwendet werden. Ebenso stellen sich Fragen nach der Verantwortung, die Carlson mit dem Begriff der „algorithmic authority“ und Gillespie mit „algorithmic objectivity“ umschreibt.³⁸ Wer ist für die Kontrolle der Daten zuständig? Der Anbieter der Software, der Kunde oder die (Medien-) Organisation, der die Daten verwendet und kennzeichnen muss, oder der Drittanbieter, der Daten für eine journalistische Verwertung zur Verfügung stellt? Die ethische Diskussion, wie Young und Hermida feststellen, „how decisions of inclusion and exclusion are made, what styles of reasoning are employed, whose values are embedded into the technology, and how they affect public understanding of complex issues“³⁹, muss besonders für den algorithmischen Journalismus geführt werden.

Das Genre stellt damit sowohl Anbieter als auch Medienorganisationen vor schwierige Herausforderungen, wie sie mit ethischen Fragestellungen zu journalistischen Werten, Normen und Transparenz umzugehen haben. Die technischen Möglichkeiten und Limitierungen der Textgenerierung und erste Studien zum Vertrauen und zur Qualität in automatisierte Inhalte zeigen Bereiche für eine Integration des algorithmischen Journalismus in den medialen Produktionsprozess auf. Wenn Anbieter und Kunden sich ihrer journalistischen Verantwortung gegenüber der Vermittlung von Öffentlichkeit bewusst sind und gleichzeitig ihre Werte und Normen transparent kommunizieren und in den Programmcode der Produkte integrieren, dann scheint es möglich, dass auch Algorithmic Journalism – im Sinn der skizzierten institutionalistischen Sichtweise – journalistische Aufgaben und Werte wahrnehmen, transportieren und kommunizieren kann.

Quellen

Anderson, C.W. (2012): Towards a sociology of computational and algorithmic journalism, in: *New Media & Society*, Nr. 7, Vol. 15, S. 1005-1021.

³⁷ Vgl. van Dalen (2012), S. 648.

³⁸ Vgl. Carlson (2014); Gillespie (2014).

³⁹ Vgl. Young und Hermida (2014), S. 4.

- Anderson, C.W. (2013): What aggregators do: Towards a networked concept of journalistic expertise in the digital age, in: *Journalism*, Nr. 8, Vol. 14, S. 1008-1032.
- Broussard, M. (2014): Artificial Intelligence for Investigative Reporting. Using an expert system to enhance journalists' ability to discover original public affairs stories, in: *Digital Journalism* (online first), DOI: 10.1080/21670811.2014.985497.
- Carlson, M. (2014): The Robotic Reporter. Automated journalism and the redefinition of labor, compositional forms, and journalistic authority, in: *Digital Journalism* (online first), DOI: 10.1080/21670811.2014.976412.
- Carstensen, K.-U.; Ebert, Ch.; Ebert, C.; Jekat, S.; Langer, H.; Klabunde, R. (Hrsg.)(2010): *Computerlinguistik und Sprachtechnologie. Eine Einführung*, 3. Auflage, Heidelberg.
- Clerwall, Christer (2014): Enter the Robot Journalist. Users' perceptions of automated content, in: *Journalism Practice*, Nr. 5, Vol. 8, S. 519-531.
- Coddington, M. (2014): Clarifying Journalism's Quantitative Turn. A typology for evaluating data journalism, computational journalism, and computer-assisted reporting, in: *Digital Journalism* (online first), DOI: 10.1080/21670811.2014.976400.
- Diakopoulos, N. (2014): Algorithmic Accountability. Journalistic investigation of computational power structures, in: *Digital Journalism* (online first), DOI: 10.1080/21670811.2014.976411.
- Gillespie, T. (2014): The Relevance of Algorithms, in: Gillespie, T.; Boczkowski, P.; Kirsten, F. (Hrsg.): *Media Technologies. Paths Forward in Social Research*, London, S. 167-194.
- Gynnild, A. (2014): Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets, in: *Journalism*, Nr. 3, Vol. 15, S. 713-730.
- Gray, J.; Bounegru, L.; Chambers, L. (2012): *The Data Journalism Handbook. How Journalists Can Use Data to Improve the News*, Sebastopol.
- Hamilton, J.T.; Turner, F. (2009): *Accountability Through Algorithm. Developing the Field of Computational Journalism. A report from Developing the Field of Computational Journalism, a Center For Advanced Study in the Behavioral Sciences Summer Workshop*, July 27-31.
- Kiefer, M-L. (2001): *Medienökonomik*, München.
- Latzer, M.; Hollnbuchner, K.; Just, N.; Saurwein, F. (2014): The economics of algorithmic selection on the Internet. Working Paper – Media Change & Innovation Division. University of Zurich, Zurich, Online:

- http://www.mediachange.ch/media//pdf/publications/Economics_of_algorithmic_selection_WP.pdf.
- Lewis, S.C.; Westlund, O. (2014): Big Data and Journalism. Epistemology, expertise, economics, and ethics, in: Digital Journalism (online first), DOI: 10.1080/21670811.2014.976418.
- Malik, M. (2004): Journalismusjournalismus. Funktion, Strukturen und Strategien der journalistischen Selbstthematization, Wiesbaden.
- Meier, K. (2011): Journalistik, Konstanz.
- Manovich, L. (2013): Software takes command, London.
- Mayer-Schönberger, V.; Cukier, K. (2013): Big Data. A Revolution That Will Transform How We Live, Work, and Think, New York.
- Mayo, J.; Leshner, G. (2000): Assessing the credibility of computer-assisted reporting. In: News-Paper Research Journal, Nr. 4, Vol. 21, S. 68-82.
- Napoli, P.M. (2014). Automated Media: An Institutional Theory Perspective on Algorithmic Media Production and Consumption, in: Communication Theory, Nr. 3, Vol. 24, S. 340-360.
- Neuburger, C. (2002): Vom Unsichtbarwerden des Journalismus im Internet, in: Hohlfeld, R.; Meier, K.; Neuburger, C. (Hrsg.): Innovationen im Journalismus. Forschung für die Praxis, Münster, S. 25-68.
- Örnebring, H. (2010): Technology and Journalism-as-labour: historical perspectives, in: Journalism, Nr. 1, Vol. 11, S. 57-74.
- Pavlik, J. (2000): The impact of technology on journalism, in: Journalism Studies, Nr. 2, Vol. 1, S. 229-237.
- Pavlik, J. (2013): Innovation and the Future of Journalism, in: Digital Journalism, Nr. 2, Vol. 1, S. 181-193.
- Rieder, B. (2014): Engines of Order: Social Media and the Rise of Algorithmic Knowing. Paper presented at the Social Media and the Transformation of Public Space conference, University of Amsterdam, June 18-20.
- Reiter, E./Dale, R. (2000): Building Natural Language Generation Systems, Cambridge.
- Van Dalen, A. (2012): The algorithms behind the headlines. How machine-written news redefines the core skills of human journalists, in: Journalism Practice, Nr. 5-6, Vol. 6, S. 648-658.
- Van der Kaa, H.; Kraemer, E. (2014): Journalist versus news consumer: The perceived credibility of machine written news, in: Research Paper presented at the 2014 Computation + Journalism Symposium, Columbia University, New York. Online: http://compute-cuj.org/cj-2014/cj2014_session4_paper2.pdf.

Wallace, J.; Dörr, K. (2014): Beyond Traditional Gatekeeping. How Algorithms and Users Restructure the Online Gatekeeping Process, Conference Paper, Digital Disruption to Journalism and Mass Communication Theory, 2-3 October 2014, Brussels.

Young, M.L.; Hermida, A. (2014): From Mr. and Mrs. Outlier To Central Tendencies, in: Digital Journalism (online first), DOI: 10.1080/21670811.2014.976409.