

Christina Elmer / Lorenz Matzat

Handbuch Daten und KI im Journalismus

HERBERT VON HALEM VERLAG

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie; detaillierte
bibliografische Daten sind im Internet über
<https://dnb.de> abrufbar.

Christina Elmer / Lorenz Matzat
Handbuch Daten und KI im Journalismus
Praktischer Journalismus, 112
Köln: Herbert von Halem Verlag 2024

© 2024 by Herbert von Halem Verlag, Köln

Alle Rechte, insbesondere das Recht der Vervielfältigung
und Verbreitung sowie der Übersetzung, vorbehalten.
Kein Teil des Werkes darf in irgendeiner Form (durch
Fotokopie, Mikrofilm oder ein anderes Verfahren)
ohne schriftliche Genehmigung des Verlages reproduziert
oder unter Verwendung elektronischer Systeme
(inkl. Online-Netzwerken) gespeichert, verarbeitet,
vervielfältigt oder verbreitet werden.

ISSN 1617-3570

ISBN (Print) 978-3-7445-2102-4
ISBN (PDF) 978-3-7445-2103-1

Den Herbert von Halem Verlag erreichen Sie auch im
Internet unter <https://www.halem-verlag.de>
E-Mail: info@halem-verlag.de

Lektorat: Imke Hirschmann
Druck: Druckhaus Nord GmbH
Satz: Herbert von Halem Verlag
Gestaltung: Claudia Ott Grafischer Entwurf, Düsseldorf

Inhaltsverzeichnis

- 9 **Vorwort**

- 11 **Einführung: Daten und Künstliche Intelligenz im Journalismus**
 von Christina Elmer und Lorenz Matzat

- 26 **Quellen und offene Daten im Journalismus**
 von Stefan Wehrmeyer

- 40 **Lokaler und regionaler Datenjournalismus**
 von Jan Georg Plavec

- 61 **Investigation auf der Basis von Leaks und großen Datensätzen**
 von Sophia Baumann und Hannes Munzinger

- 77 **Visualisierung: Grundlagen und Rezeption datenbasierter
 Grafiken**
 von Gianna-Carina Grün

- 97 **Automation im Journalismus**
 von Timo Grossenbacher

- 116 **Geodaten und Satellitendaten im journalistischen Einsatz**
 von Achim Tack

- 140 **Datenbasierter Journalismus zur Klimakrise**
 von Sören Müller-Hansen

- 159 **Interaktiver Datenjournalismus: Dynamische Visualisierung, Tools und Beteiligung**
von Hendrik Lehmann
- 182 **Datenjournalismus und Dashboards in der COVID-Pandemie**
von Marie Gundlach
- 199 **Der Kampf um Wasser: Wie wir mit lokalen Daten den Klimawandel greifbar machen**
von Max Donheiser und Katarina Huth
- 216 **Ethik des datenbasierten Journalismus**
von Jessica Heesen
- 232 **Daten in der Rezeptionsanalyse: Ansätze und Methoden der Editorial Intelligence**
von Fabian Maier und Ana Moya
- 249 **Datenjournalismus aus Sicht der Kommunikationswissenschaft**
von Mario Haim und Valerie Hase
- 265 **Methoden und Strategien im Algorithmic Accountability Reporting**
von Nicolas Kayser-Bril
- 282 **KI im Bayerischen Rundfunk: Neue journalistische Angebote durch Automatisierung und KI**
von Uli Köppen
- 297 **KI-Assistenz in der Programmierung und Datenarbeit**
von Claus Hesseling
- 308 **Modularer (Daten-)Journalismus**
von Meik Bittkowski
- 323 **Große Sprachmodelle im Journalismus**
von Steffen Kühne

- 343 **Praktische Strategien für den KI-Einsatz im Journalismus**
von David Caswell
- 358 **The medium is the machine, oder: Die Wirklichkeit der KI**
von Christoph Kucklick
- 374 **Autor:innenverzeichnis**

Einführung: Daten und Künstliche Intelligenz im Journalismus

von Christina Elmer und Lorenz Matzat

Journalismus war schon immer Technologietreiber. Zahlreiche Fortentwicklungen bei den Druckmaschinen ergaben sich über die Jahrhunderte aus den Anforderungen, täglich oder sogar mehrmals am Tag Ausgaben von Zeitungen veröffentlichen zu können. Zwar waren die technischen Berufe von den redaktionellen getrennt, trafen aber etwa im Aufgabenbereich der Setzer aufeinander, die immer wieder auch sprachlich letzte Hand anlegten. Mit den Produktionen für das Radio und später das Fernsehen verschmolzen technische und journalistische Herangehensweise noch weiter. Die ersten Schritte für Online-Ausgaben von Zeitungen und Magazinen im World Wide Web Ende der 90er-Jahre des vergangenen Jahrhunderts machten nicht selten Redakteur:innen.

Journalismus ist inzwischen aber auch getrieben und herausgefordert von Technologie (vgl. Meier et al. 2023). So erforderte das Internet mit seinen Echtzeitöffentlichkeiten eine Beschleunigung der redaktionellen Routinen: Ereignisse werden nicht mehr nur tagesaktuell eingeordnet, sondern stundenaktuell, teilweise in Livetickern. Zudem müssen sich Verlage mit immer neuen Social-Media-Plattformen auseinandersetzen: Wo können relevante Publika erreicht und welche publizistischen Formate müssen dafür entwickelt werden? Nicht zuletzt wirken sich auch die Nutzungsgewohnheiten von Technologie auf den Journalismus aus. So werden digitale Medien heute überwiegend auf Mobilgeräten rezipiert, wodurch Anpassungen in der Erzählweise und Darstellung von Inhalten notwendig wurden.

Treiber und getrieben: Diese Dichotomie prägt auch den Datenjournalismus. Wie kein anderes Feld sorgte er dafür, dass Datenbearbeitung, Softwareentwicklung und Designtätigkeiten integrierte Bestandteile redaktioneller Arbeit wurden. Zugleich sind gerade diese Tätigkeiten von Technologietrends oftmals unmittelbar betroffen. Mittels Künstlicher Intelligenz können sie bereits heute unterstützt und zum Teil sogar komplett automatisiert werden. Während sich andere Domänen mit ihrer Reaktion auf die KI-Transformation Zeit lassen konnten, waren datenjournalistische Teams längst involviert. Sie leisteten wertvolle Pionierarbeit, betreuten entsprechende Entwicklungsprojekte und waren maßgeblich an KI-Richtlinien beteiligt. In der Praxis sind Datenjournalismus und KI also bereits eng miteinander verknüpft. Deswegen widmet sich dieses Handbuch beiden Feldern gemeinsam.

Beim Begriff des Datenjournalismus ist offen, ob er primär für eine Methodik, ein Format oder ein Genre steht (vgl. Royal/Blasingame 2015). Diese Frage wird auch dieses Handbuch nicht final beantworten. Es kann lediglich Orientierung bieten, indem es unterschiedliche Perspektiven vorstellt und die jeweils zentralen Aspekte somit vergleichbar macht. In der journalistischen Praxis hat sich der Datenjournalismus abhängig vor allem von bestehenden Strukturen und konkreten Herausforderungen verschiedentlich eingefügt. Deutlich wird diese Vielfalt dadurch, dass viele Autor:innen dieses Handbuchs Teil von datenjournalistische Teams aus dem deutschsprachigen Raum sind.

Datenjournalismus – quantitative Methoden für mehr Tiefe

Am Anfang stand die Präzision. Das Buch *Precision Journalism: A Reporter's Introduction to Social Science Methods* gilt vielfach als Ausgangspunkt des heutigen Datenjournalismus. Geschrieben wurde es von Philip Meyer, der es 1973 noch als Journalist veröffentlichte und später Professor an der University of North Carolina wurde. Er plädierte dafür, sozialwissenschaftliche Methoden wie Experimente, Statistiken und Umfragen stärker in die journalistische Arbeit zu integrieren, um eine genauere und tiefergehende Berichterstattung zu erreichen (vgl. Meyer 1973). Herkömmliche Recherche und Narration wollte Meyer durch die quantitativen Methoden keinesfalls ersetzen, sondern ergänzen. Diese Sichtweise bestärkte auch

Steve Doig, Journalismusprofessor an der Arizona State University, etwa 40 Jahre später in einem Interview. Datenjournalismus sei für ihn »a social science done on deadline« (Remington 2012). Damit verweist Doig auch auf die spezifischen Rahmenbedingungen, die den Methoden aus der Wissenschaft im Journalismus gesetzt sind. Deadlines gehören ebenso dazu wie Kriterien der Themenauswahl und der Anspruch einer allgemeinverständlichen Kommunikation.

Mit dem Anspruch an einen präziseren Journalismus begann eine Transformation, die als »quantitative turn« bezeichnet wird – eine Hinwendung zu quantitativen Methoden innerhalb der journalistischen Recherche und Produktion (vgl. Coddington 2014). In den 1980er-Jahren entwickelte sich daraus im angloamerikanischen Raum das Computer Assisted Reporting (CAR). Was mit dem Begriff konkret gemeint ist, stellte der langjährige Investigativjournalist und spätere Journalismusprofessor Brant Houston klar: Computer Assisted Reporting bedeute mehr, als den Computer zum Schreiben zu verwenden oder damit im Internet zu surfen. Vielmehr umfasse der Begriff Techniken, Datenbanken zu analysieren oder selbst aufzubauen, um Kontext zu liefern und einen Anhaltspunkt für komplexe Recherchen zu finden, die ein Thema umfassender und mit einem tieferen Verständnis darstellen können (Houston 2018: 5). Der Begriff des Datenjournalismus wird häufig synonym zu Computer Assisted Reporting verwendet – vor allem im europäischen Raum, wo der Journalismus die quantitativen Methoden des Datenjournalismus mit einigen Jahren Verspätung für sich entdeckte.

Von CAR zu DDJ

Ende der 2000er-Jahre trafen mehrere Entwicklungen aufeinander. Das Aufkommen des mobilen Internets zeichnete sich ab – das iPhone wurde Ende 2006 präsentiert – und neue Webtechnologien wie HTML5 und Javascript eröffneten flexiblere Veröffentlichungsmethoden. Aus den USA drang die Idee von Open Data und Open Government auch nach Europa. Der frisch gewählte US-Präsident Barack Obama hatte als erste Amtshandlung eine Open-Government-Direktive gezeichnet, die transparentes Regierungshandeln versprach. Im britischen *Guardian* veröffentlichten 2006 zwei Redakteure den Aufruf »Give us back our crown jewels« und meinten damit die Datensätze der öffentlichen Hand (Arthur/Cross 2006).

Drei Jahre später – der *Guardian* war zu der Zeit noch nicht der internationale Medienplayer der heutigen Tage – begann dort der Redakteur Simon Rogers, nahezu täglich Datensätze aller Art zu veröffentlichen und mit damals noch kruden Datentools zu visualisieren.

Inspiziert dazu hatte Rogers nicht zuletzt die Herangehensweise von Adrian Holovaty (Gambini 2019). Der ausgebildete Journalist, damals Softwareentwickler bei einer US-amerikanischen Regionalzeitung, hatte ebenfalls 2006 einen wegweisenden Text geschrieben: *A fundamental way newspapers sites have to change*. Hier skizzierte Holovaty, wie Redaktionen ihre eigene Berichterstattung in Datensätzen herunterbrechen und öffentliche Daten, Open Data, für ihre Arbeit verwenden könnten. Ideen, die sich rund zehn Jahre später in Ansätzen wie dem »structured journalism« und generell der Automatisierung von Journalismus wiederfanden (Holovaty 2006)

2007 hatte die Deutsche Presse-Agentur (dpa) ein Datenjournalismusteam aufgebaut, das unter dem Label »Regiodata« auf regionale Verbreitungsgebiete zugeschnittene Datenauswertungen erstellte. Zu früh offenbar: Es wurde mangels Nachfrage drei Jahre später wieder eingestellt. Im gleichen Jahr, 2010, veröffentlichten internationale Medien mithilfe der Plattform Wikileaks ihre Berichterstattung zu den Afghanistan-Protokollen. Der Kopf von Wikileaks, Julian Assange, saß auch deswegen lange in Isolationshaft in Großbritannien.

Während die Projektpartner *Guardian* und *New York Times* bereits umfangreiche Datenvisualisierungen im Netz anboten, konnte der deutsche Partner *Spiegel Online* nicht annähernd mithalten. Doch in den kommenden Jahren begannen auch deutschsprachige Medien mit dem neuartigen »data-driven journalism« (ddj), der entgegen CAR Computer nicht primär zur Datenanalyse, sondern auch die Datensätze direkt für interaktive Datenvisualisierungen und Karten verwendete. *Zeit Online*, *Taz* und *Spiegel* begannen vermehrt, aufwendige Datenvisualisierungen zu veröffentlichen. 2011 wurde erstmals ein Datenjournalismusstück mit einem Grimme Online Award ausgezeichnet.

Seit dem bahnbrechenden Wikileaks-Datensatz von 2009 sollte die investigative Arbeit mit Datenleaks den Datenjournalismus begleiten. Dabei werden große Mengen an Dokumenten strukturiert ausgewertet, die den Redaktionen nicht selten über Whistleblower zur Verfügung gestellt und in internationalen Netzwerken analysiert werden. Ein prägendes Beispiel waren die Panama Papers – ein vom International Consortium of Investigative Journalists (ICIJ) koordiniertes globales Rechercheprojekt zu

Praktiken der Steuervermeidung, basierend auf zugespielten Unterlagen des Offshore-Dienstleisters Mossack Fonseca (ICIJ 2017).

Derart prominente Veröffentlichungen haben spürbar dazu beigetragen, die Nachfrage nach datenjournalistischen Kompetenzen in Redaktionen zu erhöhen. Auch im deutschsprachigen Raum bildeten Medienhäuser eigene Ressorts, verstärkten ihre investigativen Teams und schufen zusätzliche Stellen. Jedoch gelang dies vor allem größeren Medienhäusern mit starken Online-Redaktionen und einer progressiven Digitalstrategie. Denn auch wenn datenjournalistische Techniken die redaktionellen Prozesse an unterschiedlichen Stellen unterstützen und die Ergebnisse potenziell in alle journalistischen Formate einfließen können (vgl. Bradshaw 2012), werden ihre Stärken im Digitalen doch besonders deutlich. So können interaktive Karten und Infografiken nicht nur Muster zeigen, sondern es dem Publikum auch erlauben, tiefer in die verwendeten Daten einzutauchen und Bezüge zur eigenen Lebensrealität zu finden. Sie transportieren damit zentrale und personalisierbare Ergebnisse der Recherche und zugleich vertrauensbildende Signale zur verwendeten Methode. Denn wer seinen Datensatz auf diese Weise im Detail darstellt, macht Schlussfolgerungen nachvollziehbar und den Prozess der Recherche damit transparent.

Visualisierung und Interaktivität als charakteristische Merkmale des Datenjournalismus – so beschreibt es auch Lorenz Lorenz-Meyer, langjähriger Online-Journalismus-Professor an der Hochschule Darmstadt (Lorenz-Meyer 2011). Jedoch klassifiziert er sie als optional, ebenso wie die Veröffentlichung des Quellenmaterials, mit dem weitere Fragestellungen exploriert werden können. Auch in einer Definition von Wiebke Loosen und Julius Reimer taucht dieser »open-data and open-source«-Ansatz als wichtiges Qualitätsmerkmal des Datenjournalismus auf (Loosen/Reimer 2017). Die somit praktizierte Offenheit bezieht sich auch auf die Recherche selbst, in die das Publikum einbezogen werden kann (Coddington 2014). Mithilfe des Crowdsourcings werden Datenanalysen realisiert, für die ursprünglich keine strukturierten Informationen vorliegen. Personen außerhalb der Redaktion können solche Projekte auf unterschiedlichen Wegen ermöglichen: Liegen ihnen einzelne der benötigten Datenpunkte privat vor, können sie diese systematisch spenden. Müssen sie erst noch erhoben werden, können diese Personen auch Messungen durchführen oder bei der Bereinigung von Datensätzen helfen, wie es in Citizen-Science-Projekten bereits etabliert ist.

Grundlegende Aspekte des datenjournalistischen Arbeitens vertiefen in diesem Handbuch die Beiträge zu Datenquellen und offenen Daten, zu

Geodaten und Satellitendaten, zu lokalem und regionalem Datenjournalismus, zur Visualisierung und zur Interaktivität in datenjournalistischen Formaten. Ein weiterer Beitrag beleuchtet die spezifischen Anforderungen der Investigation mit Datenleaks und der Arbeit in Recherchenetzwerken. Inhaltlich werden zudem die Rolle von Daten beim Journalismus zur Klimakrise und ein Rechercheprojekt zum Wasserverbrauch beispielhaft vorgestellt. Ein Beitrag zeigt die kommunikationswissenschaftliche Sicht auf den Datenjournalismus, ein weiterer diskutiert ethische Fragen rund um den Einsatz von Daten, Algorithmen und Künstlicher Intelligenz im Journalismus.

Automatisierung – Algorithmen im Dienst der Redaktion

Wer in Recherchen mit großen Datenmengen arbeitet, benötigt zwingend spezifische Werkzeuge. Schon bei der Bereinigung von Datensätzen wiederholen sich viele Arbeitsschritte, Gleiches gilt für die Analyse und die Aufbereitung in Texten und Infografiken. Da sich datenjournalistische Teams häufig mit wechselnden Themen befassen, bilden sich dabei nur schwer Routinen heraus, wohl aber wiederkehrende Praktiken (Ausserhofer 2015). Dabei gilt: Was sich regelbasiert wiederholt, kann mithilfe von Code automatisiert werden. Effizienter Datenjournalismus bedeutet, die Gelegenheiten für Automatisierung zu erkennen und mithilfe bestehender Werkzeuge oder Programmierung zu lösen. Inzwischen stehen allerdings nicht mehr nur Datenjournalist:innen vor dieser Herausforderung. Mithilfe lernender Systeme und entsprechender Software lassen sich zahlreiche Aufgaben entlang der gesamten redaktionellen Wertschöpfungskette automatisieren (Diakopoulos 2019). Vom Monitoring aktueller Daten über das regelbasierte Texten von Artikeln bis zur Optimierung von Überschriften für Suchmaschinen – Automatisierung durch Software ist selbstverständlicher Bestandteil redaktioneller Prozesse, ohne dabei menschliche Arbeit überflüssig zu machen. Sie wird weiterhin gebraucht, um die Automatisierung zu gestalten und die journalistischen Aufgaben zu betreuen, die Maschinen nicht übernehmen können. Der Stellenabbau in manchen Verlagen in der jüngsten Zeit lässt allerdings befürchten, dass Automatisierung niedrig qualifizierte Jobs durchaus überflüssig macht oder zumindest als Argument für Einsparungen vorgeschoben wird.

Konkret lässt sich der Wandel der Arbeit am Beispiel des sogenannten »Roboterjournalismus« nachvollziehen. Im Kern geht es dabei um die automatische Generierung journalistischer Beiträge, vor allem von Texten, Infografiken und Karten. Die Basis einer solchen Automatisierung bilden Datensätze mit vorhersehbarer Struktur, etwa zu Börsenkursen oder zum Wetter, zu Sportereignissen oder Wahlergebnissen. Diese Daten werden weitgehend regelbasiert in vorher festgelegte Module übersetzt, wobei Künstliche Intelligenz bislang aus Gründen der Qualitätssicherung noch selten verwendet wird. Doch das muss nicht so bleiben. Hypothetisch könnten fortgeschrittene KI-Modelle den beschriebenen Prozess schon bald komplett übernehmen – eine Zukunftsvision, die angesichts des Tempos der aktuellen Entwicklungen heute durchaus realistisch erscheint. Mario Haim und Andreas Graefe (2022) beschreiben drei wesentliche Ziele, die Medienhäuser mithilfe von Automatisierung verfolgen. Erstens ist das die Skalierung der Berichterstattung, indem mannigfaltige Versionen einer Story erstellt werden, die jeweils speziell auf die Interessen bestimmter Zielgruppen zugeschnitten sind. Zweitens geht es um eine inhaltliche Erweiterung, wenn mithilfe automatisierter Workflows über Themen berichtet werden kann, für die ansonsten die Ressourcen fehlen würden. Drittens werde auch automatisiert, um die technischen Möglichkeiten zu demonstrieren – ein Phänomen, das insbesondere rund um den Launch von ChatGPT zu beobachten war.

Während der COVID-19-Pandemie hat sich auf diese Weise ein neuartiges datenjournalistisches Format herausgebildet: In Dashboards – ähnlich einem Armaturenbrett mit einem Set verschiedener kleinerer Datenvisualisierungen – wurden zunächst aktuelle Inzidenzen dargestellt, später die Klinikauslastung. Das besondere Merkmal dieser Dashboards war gar nicht primär die Automatisierung, sondern vielmehr ihre Prominenz und der damit zusammenhängende Anspruch: Als feste Elemente auf den Startseiten von Online-Medien standen sie permanent im Fokus der Aufmerksamkeit und mussten intensiv gepflegt werden. Überhaupt waren datenjournalistische Analysen während der Pandemie enorm gefragt, boten sie doch die Möglichkeit einer evidenzbasierten Einordnung des von Unsicherheiten geprägten Geschehens. Das Standing der Datenteams in den Redaktionen wurde dadurch gestärkt, zugleich haben sich ihre Praktiken gewandelt. Dashboards und andere automatisierte Formate haben sich etabliert, folglich ist die Sichtbarkeit datenjournalistischer Arbeit in der aktuellen Berichterstattung gewachsen. Skalierungseffekte

werden dadurch möglich, Spielräume hingegen kleiner. Nach der Hochphase der Pandemie folgten andere Dashboards zu aktuellen Themen wie Energieversorgung oder Klimakrise.

In diesem Handbuch beschäftigen sich zwei Beiträge dezidiert mit diesen Aspekten – der Automatisierung im Datenjournalismus und den datenbasierten Dashboards in der COVID-19-Pandemie. Ein weiterer Beitrag schlägt die Brücke zur statistischen Analyse von Nutzungsdaten im digitalen Journalismus. Auch um redaktionelle Entscheidungsprozesse zu unterstützen, werden automatisierte Verfahren und fortgeschrittene Analysetechniken angewandt. Betrachtet man den Umgang mit Daten in Medienhäusern, darf dieser Bereich nicht fehlen, zumal er aus methodischer Sicht besonders gehaltvoll ist.

Lernende Systeme – KI-Werkzeuge als Sparringspartner

Das Thema Künstliche Intelligenz (KI) führt ein Doppelleben in den Redaktionen. Es ist zugleich Hype und Alltagswerkzeug, also bereits ständig präsent und zugleich mit deutlich überhöhten Erwartungen verbunden. Schließlich versammeln sich unter diesem Begriff technologische Werkzeuge unterschiedlichster Komplexität, von einfacheren Prognose-Algorithmen bis zu generativen Deep-Learning-Modellen, deren Funktionsweise nicht im Detail nachvollzogen werden kann. Sie alle haben aber das Grundprinzip gemeinsam. Künstliche Intelligenz umfasst algorithmische, selbstlernende Systeme, die aus großen Datenmengen Muster ableiten und diese auf neue Daten anwenden können. Auf diese Weise können sie lernen, Objekte auf Bildern zu unterscheiden, Verhaltensspuren zu klassifizieren oder plausible Texte zu generieren – natürlich nicht fehlerfrei, aber inzwischen auf einem Niveau, das noch vor wenigen Jahren unvorstellbar war. Das änderte sich Ende 2022. Im November präsentierte das Unternehmen OpenAI seinen Chatbot ChatGPT, der auf einem umfangreich trainierten Sprachmodell aufbaut. Dieses und weitere Modelle belegten eindrucksvoll: Algorithmische Systeme können inzwischen auch Sprache verarbeiten und Konversationen führen, die durchaus plausibel und sinnhaft wirken. Im Folgejahr 2023 wurde der Umgang mit KI-Systemen breit diskutiert – sowohl gesamtgesellschaftlich als auch in Redaktionen und journalistischen Fachmedien.

Obwohl wir noch am Anfang der KI-Transformation stehen, werden lernende Systeme schon heute vielfach in den Medien eingesetzt (New-

man 2023; Beckett/Yaseen 2023). Sie unterstützen Wissensmanagement und Recherche, Produktion und Distribution. Dabei ist wichtig zu verstehen: Es ist mitnichten ein einziges KI-System, das hinter den diversen Anwendungen in der journalistischen Wertschöpfungskette steckt. Stattdessen sind es viele unterschiedliche Systeme, deren Modelle jeweils für eine bestimmte Aufgabe trainiert wurden – und die auch nur für diese Aufgabe eingesetzt werden können. Künstliche Intelligenz ist also keine umfassende technologische Macht im Hintergrund dieser Anwendungen, sondern beschreibt nicht mehr als eine gemeinsame Fähigkeit – zumindest bislang. Wenn also über »die KI« gesprochen wird, erwecken wir allzu schnell einen falschen Eindruck. Besser wäre es derzeit, konkreter zu beschreiben, mit welchem System man es zu tun hat und wofür es trainiert wurde. Ein weiteres Missverständnis hat uns das frühe, allgemein zugängliche ChatGPT von Ende 2022 beschert. Derartige Sprachmodelle sind primär auf Plausibilität trainiert, worunter der Wahrheitsgehalt der Antworten leidet – die Modelle erzeugen falsche Aussagen, halluzinieren oder konfabulieren. Inwieweit diese Halluzinationen Kinderkrankheiten einer recht neuen Technologie sind und absehbar eingehegt werden können, ist derzeit noch unklar. Schon jetzt sind Sprachmodelle nicht gänzlich ungeeignet für den Einsatz im Journalismus. Brauchbare KI-Assistenten sind möglich, wenn lernende Algorithmen als Team arbeiten: Ein Sprachmodell kann mit einem weiteren KI-System verknüpft werden, das an eine Datenbank mit gesicherten Fakten angeschlossen ist und nachvollziehbare Quellen liefert. Erste prototypische Anwendungen demonstrieren dieses Potenzial.

Wo assistieren lernende Systeme heute in den Redaktionen? Im Bereich des Wissensmanagements helfen sie bereits dabei, dass digitale Inhalte in Archivbeständen besser gefunden werden können. So digitalisierte der Westdeutsche Rundfunk sein Archiv mithilfe von KI-Systemen, die gesprochenes Wort in Text umwandeln oder Personen automatisch erkennen. Eine zielgerichtete Suche nach bestimmten Aussagen oder Szenen ist somit möglich (Maroni et al. 2020). Auch Sammlungen von Textbeiträgen, die zuvor händisch verschlagwortet wurden, können mit KI-Unterstützung strukturiert und automatisch Themen zugeordnet werden. Die beschriebenen Anwendungen lassen erahnen, welches Potenzial in diesem Bereich noch schlummert: Würden auch Materialien aus der Recherche in derartige Systeme aufgenommen, könnte das Wissen einer Redaktion deutlich wirksamer als heute geteilt und über Schreibtischgrenzen hin-

weg wirken. Sogenannte Enterprise-Search-Systeme könnten die redaktionelle Zusammenarbeit umwälzen, sofern Materialien offen geteilt würden und ein geschützter Datenaustausch möglich wäre.

Multi-modal

Auch für Recherchen verwenden Redaktionen bereits KI-Technologien. Werden besonders große Datensätze oder Dokumentenberge ausgewertet, sind häufig lernende Systeme beteiligt. Wie schon beschrieben, strukturieren sie Leaks in investigativen Recherchen, um Netzwerke aus Akteuren auszuleuchten und verborgene Strukturen sichtbar zu machen. Im Datenjournalismus werden KI-Systeme teilweise sogar selbst trainiert, etwa um Muster in Satellitendaten zu finden oder Transponderdaten von Flugzeugen in großem Maßstab zu untersuchen. Die zunehmende Datafizierung des gesellschaftlichen Lebens dürfte noch viele weitere Ansätze für derartige Recherchen liefern. Denn immer mehr Vorgänge hinterlassen ohnehin Datenspuren, in denen sich Strukturen und Verbindungen identifizieren lassen. Und wo derartige Spuren fehlen, erheben Redaktionen mithilfe von Datenspenden oder Sensoren mitunter eigene Daten. KI-Technologien können dabei unterstützen, in diesen oftmals sehr unstrukturierten Daten Muster zu erkennen und damit Hinweise für weitere Recherchen zu entdecken.

Darüber hinaus ist die neue Klasse »multi-modaler« Modelle vielversprechend: Sie können neben Text auch Audio, Bilder und Bewegtbilder sowohl generieren als auch auswerten und kontextualisieren: Nimmermüde Augen und Ohren, die zahlloses Video- und Tonmaterial, live monitoren können. Allerdings können die gleichen KI-Systeme die journalistische Recherche auch erschweren. Denn generative Modelle werden zunehmend von Akteuren eingesetzt, die mithilfe synthetischer Inhalte bewusst Desinformationen erzeugen und verbreiten. KI-generierte Bilder, Tonaufnahmen und Videos, sogenannte Deepfakes, sind oftmals mit bloßem Auge kaum zu erkennen. Daher werden bereits KI-Systeme entwickelt, die Deepfakes zum Beispiel anhand ihrer Datenstruktur erkennen sollen – ein technologisches Wettrennen, das wohl kaum zu gewinnen sein wird. Und dennoch ist es für Journalist:innen wichtig, derart unterstützende Werkzeuge für die Verifikation zur Verfügung zu haben. Schließlich können sie durchaus entscheidende Hinweise geben – zumal

sich viele Deepfakes bei genauerem Hinsehen als Cheapfakes entpuppen und überraschend einfach zu identifizieren sind.

Produktion

Ist die Recherche abgeschlossen, wird innerhalb der Produktion aus dem gesammelten Material ein journalistischer Beitrag. Dabei stellt sich unter anderem die Herausforderung, unterschiedliche Medienformate ineinander umzuwandeln: Videobeiträge benötigen Untertitel, aufgezeichnete Interviews müssen transkribiert, Artikel automatisiert vorgelesen werden. Für diese Aufgaben verwenden Redaktionen bereits vielfach KI-Technologien aus den Bereichen Text-to-Speech und Speech-to-Text. Auch wird bereits mit synthetischen Stimmen experimentiert, die realen Autorinnen und Autoren nachgebildet sind. Dahinter steht der Gedanke, mit journalistischen Angeboten noch besser auf die Gewohnheiten und Bedürfnisse der Rezipierenden einzugehen. Mithilfe von KI-Systemen könnte dieses Prinzip noch deutlich weiterentwickelt werden: Inhalte würden automatisch übersetzt, persönliche Interessen, Wissensstände und Nutzungssituationen ließen sich bei Präsentation und Stil berücksichtigen. Natürlich müsste eine derartige Personalisierung verständlich erklärt werden, und es wäre wichtig zu verhindern, dass Menschen letztlich schlechter über relevante Ereignisse informiert sind. Derartige Fragen zu den medienethischen Dimensionen der Arbeit mit lernenden Systemen haben sich auch viele Redaktionen bereits gestellt. Antworten finden sich unter anderem in KI-Guidelines, in denen sich Medien unter anderem zu Maßnahmen der Qualitätssicherung sowie Transparenz in automatisierten Prozessen verpflichten (Cools 2023).

Ein Text des Handbuchs widmet sich der Idee des »structured journalism« oder »modular journalism«, der Personalisierung gangbar macht. Redaktionelle Formate könnten in vorher definierte Bausteine zerlegt werden, die je nach Nutzungsbedürfnis und Situation passend und automatisiert zu kombinieren wären – ein Konzept, das seit einigen Jahren exploriert wird und in den Bereich der Automatisierung fällt. Jedoch müsste sich der digitale Journalismus dafür in mehreren Dimensionen grundlegend weiterentwickeln: Technisch würden neuartige Content-Management-Systeme mit umfangreicheren Metadaten benötigt. Und publizistisch müssten auch die Geschichten entsprechend modular erzählt werden – nicht mit einem einzigen roten Faden, sondern vielen möglichen Optionen.

Veröffentlichung

Welche Rolle multi-modale Modelle bei der Veröffentlichung journalistischer Inhalte spielen werden, ist noch nicht absehbar. Sie haben das Potenzial, direkt personalisierte Daten- und Kartenvisualisierungen sowie Audio- und Videobeiträge zu erstellen. Das wirft Fragen auf: Inwieweit werden sich journalistische Qualitätsansprüche behaupten lassen, wenn diese Personalisierung nicht (nur) aufseiten der Kanäle journalistischer Verlage geschieht, sondern auf den Geräten der Nutzer:innen durch die Systeme der großen Technologiekonzerne selbst? Eine mögliche Entwicklung durch KI sind Assistenzsysteme auf Betriebssystemebene der Mobilgeräte, die das Internet permanent im Auftrag der Nutzer:innen nach interessanten Inhalten durchsuchen und die diese Inhalte »vorkauen« und dann nach Gusto der Nutzer:innen im Rahmen ihrer Bedürfnisse präsentieren. Vermutlich werden neue interaktive Formate entstehen, die sich mit den klassischen linearen Formaten des heutigen Journalismus, die aus Zeiten der Druckerpresse, des Radios und TVs stammen, kaum vertragen. Was kommt dann noch vom Original-Journalismus an, wenn eine neue Klasse von Gatekeepern und Formaterzeugern durch solche Assistenzsysteme entsteht?

Bislang ließen sich die im digitalen Raum publizierten journalistischen Inhalte anhand ihrer Datenspuren mithilfe von KI-Systemen monitoren. Die bereits angesprochene Analyse von Nutzungsdaten hat sich in Medienhäusern inzwischen stark ausdifferenziert und nutzt auch fortgeschrittene Methoden. Es wird nicht mehr nur verglichen, welche Beiträge auf welchen Plattformen wie häufig angeklickt und wie intensiv rezipiert werden. Es geht auch um Maßnahmen, die das jeweilige Geschäftsmodell stärken: So sollen Gelegenheitsnutzende für ein Abonnement gewonnen, Paywalls entsprechend sinnvoll ausgesteuert oder Preisanreize wirkungsvoll gesetzt werden. Werbebasierte Geschäftsmodelle koppelten bislang die Daten mit der entsprechenden Ausspielung von Anzeigen. Auch in diesem Bereich stellen sich Fragen angesichts künftig vorstellbarer KI-Anwendungen: Welche Zukunft haben Paywall- und Werbe-Geschäftsmodelle, wenn die Nutzer:innen gar nicht mehr auf die Website, in die Apps der Verlage kommen, sondern nur noch ihre KI-Assistenz vorbeischicken?

Der Blick auf die Anwendungsfelder zeigt: Die KI-Transformation fordert Journalist:innen und Redaktionen auf vielen Ebenen heraus, publizistisch und technologisch, medienethisch und im Hinblick auf Kompetenzen und Qualifikationen. Dabei ist auch die Haltung relevant: Schließlich geht es für

einen verantwortungsvollen Journalismus nicht nur darum, den KI-bedingten Wandel für die eigenen Ziele zu nutzen, sondern auch darum, die zugrunde liegenden Prämissen und gesellschaftlichen Folgen zu hinterfragen. Bei aller Experimentierfreude darf in Redaktionen also auch die kritische Distanz nicht fehlen – und die Ambiguitätstoleranz, um beides zeitgleich zu forcieren. Das betrifft auch das Verhältnis zwischen Medienhäusern und Technologiefirmen, deren technologische Werkzeuge und Systeme auch im Journalismus breit genutzt werden, wodurch Abhängigkeiten und enge ökonomische Verflechtungen entstehen können. Die medienethisch gebotene Unabhängigkeit zu wahren, gehört für den Journalismus daher zu den wesentlichen Herausforderungen dieser Zeit (Simon 2023).

In diesem Handbuch verdeutlichen mehrere Beiträge, wie lernende Systeme die journalistische Arbeit bereits heute verändern. Einer diskutiert die Nutzung großer Sprachmodelle in der redaktionellen Praxis, ein anderer beleuchtet die Assistenz durch KI-Systeme beim Programmieren und in der Arbeit mit großen Datensätzen. Wie neue journalistische Angebote und Produkte auf KI-Basis entstehen können, beschreibt ein weiterer Beitrag. Zudem geht es im Text zu Algorithmic Accountability Reporting um Recherchen, die gesellschaftlich relevante KI-Systeme kritisch hinterfragen. Schließlich wagt ein Beitrag eine Prognose hinsichtlich der redaktionellen Zukunft mit Daten und Künstlicher Intelligenz, bevor eine Reflexion zu den Auswirkungen der KI-Transformation auf die Konzepte von Wahrhaftigkeit, Information und Medien das Handbuch abschließt.

Quellen

- Arthur, Charles; Cross, Michael: Give us back our crown jewels. In: *The Guardian*, 2006. <https://www.theguardian.com/technology/2006/mar/09/education.epublic> (Abruf am 23.2.2024)
- Ausserhofer, Julian: »Die Methode liegt im Code«: Routinen und digitale Methoden im Datenjournalismus. In: Maireder, Axel; Ausserhofer, Julian; Schumann, Christina; Taddicken, Monika (Hrsg.): *Digitale Methoden in der Kommunikationswissenschaft*. Berlin [Böhlmann & Schremmer Verlag] 2015, S. 87-111. <http://dx.doi.org/10.17174/dcr.v2.5>
- Beckett, Charlie; Yaseen, Mira: *Generating Change. A global survey of what news organisations are doing with AI*. 2023. London School of Economics.

- <https://www.journalismai.info/research/2023-generating-change> (Abruf am 23.2.2024)
- Bradshaw, Paul: What Is Data Journalism? In: Bounegru, Liliana; Gray, Jonathan (Hrsg.): *The Data Journalism Handbook. Towards a Critical Data Practice*. 2012. <https://datajournalism.com/read/handbook/one/introduction/what-is-data-journalism> (Abruf am 23.2.2024)
- Coddington, Mark: Clarifying Journalism's Quantitative Turn. In: *Digital Journalism*, 3(3), 2014, 331-248. <https://doi.org/10.1080/21670811.2014.976400>
- Cools, Hannes: *Towards Guidelines for Guidelines on the Use of Generative AI in Newsrooms*. 2023. <https://generative-ai-newsroom.com/towards-guidelines-for-guidelines-on-the-use-of-generative-ai-in-newsrooms-55b0c2c1d960> (Abruf am 23.2.2024)
- Diakopoulos, Nicholas: *Automating the News. How Algorithms Are Rewriting the Media*. Cambridge [Harvard University Press] 2019. <https://doi.org/10.4159/9780674239302>
- Gambini, Letizia: A decade of working in datajournalism: what has changed? In: *medium.com*, 2019. <https://medium.com/we-are-the-european-journalism-centre/a-decade-of-working-in-data-journalism-what-has-changed-8d950d99935e> (Abruf am 23.2.2024)
- Haim, Mario; Graefe, Andreas: Automatisierung und Journalismus. In: Löffelholz, Martin; Rothenberger, Liane (Hrsg.): *Handbuch Journalismustheorien*. Wiesbaden [Springer VS] 2022. https://doi.org/10.1007/978-3-658-32153-6_28-1
- Holovaty, Adrian: A fundamental way newspapers have to change. In: *holovaty.com*, 2006. <https://www.holovaty.com/writing/fundamental-change/> (Abruf am 23.2.2024)
- Houston, Brant: *Computer-assisted Reporting. A Practical Guide*. New York [Routledge] 1996
- Houston, Brant: *Data for Journalists. A Practical Guide for Computer-Assisted Reporting*. London [Routledge] 2018
- ICIJ: *Explore the Panama Papers Key Figures*. 2017. <https://www.icij.org/investigations/panama-papers/explore-panama-papers-key-figures/> (Abruf am 23.2.2024)
- Loosen, Wiebke; Reimer, Julius: Data Journalism at Its Finest: A Longitudinal Analysis of the Characteristics of Award-Nominated Data Journalism Projects. In: Nguyen, An (Hrsg.): *News, Numbers and Public Opinion in a Data-Driven World*. New York [Bloomsbury] 2017

- Lorenz-Meyer, Lorenz: Was ist Datenjournalismus? In: *Scarlatti.de*, 2011. <http://scarlatti.de/?p=861> (Abruf am 23.2.2024)
- Maroni, Dirk; Köhler, Joachim; Fisseler, Jens; Becker, Sven: *Die ARD Mining-Plattform – Künstliche Intelligenz im produktiven Einsatz für die automatisierte Erschließung im multimedialen Produktionsprozess*. FKT – Fernseh- und Kinotechnik, 5/2020
- Matzat, Lorenz: 10 Jahre Datenjournalismus: Gemischte Gefühle. In: Krone, Jan (Hrsg.): *Medienwandel kompakt 2017-2019. Schlaglichter der Veränderung in Kommunikation, Medienwirtschaft, Medienpolitik und Medienrecht - Ausgewählte Netzveröffentlichungen*. Wiesbaden [Springer Fachmedien] 2019. <https://doi.org/10.1007/978-3-658-27319-4>
- Meier, Klaus; Graßl, Michael; García-Avilés, Jose Alberto; Mondejar, Dámaso; Kaltenbrunner, Andy; Lugschitz, Renée; Porlezza, Colin; Mazzoni, Petra; Wyss, Vinzenz; Saner, Mirco: Innovations in Journalism as Complex Interplay: Supportive and Obstructive Factors in International Comparison. In: *Media and Communication*, 12, 2023. <https://doi.org/10.17645/mac.7443>
- Meyer, Philip: *Precision Journalism*. Bloomington/IL [Indiana University Press] 1973.
- Newman, Nic: *Journalism, Media, and Technology Trends and Predictions 2023*. Reuters Institute Report. Oxford. 2023. <https://reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2023> (Abruf am 23.2.2024)
- Remington, Alex: »Social science done on deadline«: Research chat with ASU's Steve Doig on data journalism. In: *The Journalist's Resource*. 2012. Archiv: <https://journalistsresource.org/home/research-chat-steve-doig-ddata-journalism-social-science-deadline/> (Abruf am 23.2.2024)
- Rogers, Simon: Welcome to the Datablog. Looking for stats and facts? This is now the place to come. In: *The Guardian*. 2009. <https://www.theguardian.com/news/datablog/2009/mar/10/blogpost1> (Abruf am 23.2.2024)
- Royal, Cindy; Blasingame, Dale: Data journalism: An explication. In: *#ISO: the Official Research Journal on the International Symposium for Online Journalism*. 2015. <https://isojournal.wordpress.com/2015/04/15/data-journalism-an-explication/> (Abruf am 23.2.2024)
- Simon, Felix M.: Escape Me If You Can: How AI Reshapes News Organisations' Dependency on Platform Companies. In: *Digital Journalism*. 2023. <https://doi.org/10.1080/21670811.2023.2287464>