TU WIEN Informatics

Automated Journalism

The Effects of Automated News Generation on News Organizations

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Business Informatics

by

Matthias Wagner, BSc
Registration Number 01625721

to the Faculty of Informatics

at the TU Wien

Advisor: Ao. Univ.-Prof. Mag. Dr. iur. Markus Haslinger
Assistance: Priv.-Doz. Univ.Lektor Mag. Dr.phil. Stefan Weber

Vienna, 9th October, 2022

_____   _____
Matthias Wagner   Markus Haslinger

TU WIEN Informatics

Automated Journalism

The Effects of Automated News Generation on News Organizations

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Business Informatics

eingereicht von

Matthias Wagner, BSc
Matrikelnummer 01625721

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Ao. Univ.-Prof. Mag. Dr. iur. Markus Haslinger
Mitwirkung: Priv.-Doz. Univ.Lektor Mag. Dr.phil. Stefan Weber

Wien, 9. Oktober 2022

_____   _____
        Matthias Wagner   Markus Haslinger

Declaration of Authorship

Matthias Wagner, BSc

I hereby declare that I have written this work independently, have given full details of the sources and aids used, and have marked places in the work—including tables, maps and illustrations—which are taken from other works or from the Internet, either verbatim or in spirit, as borrowed, in any case indicating the source.

Vienna, 9th October, 2022

_____

Matthias Wagner

Acknowledgements

Danksagung

Ein besonderer Dank geht an alle Interviewpartner, die sich die Zeit genommen haben, mich bei meiner Arbeit zu unterstützen. Ich bin meinem Betreuer, Professor Markus Haslinger, sehr dankbar für sein großartiges, kontinuierliches Feedback, seine wertvollen Anregungen und seine großartige Unterstützung. Mein Dank gilt auch Dr. phil. Stefan Weber, der mich mit wertvollem Feedback versorgt hat. Auch geht mein Dank an Professor Christian Huemer für sein Feedback zum Proposal dieser Arbeit.

Mein aufrichtiger Dank gilt auch meinen Freunden, die sich die Zeit genommen haben, meine Arbeit Korrektur zu lesen. Abschließend möchte ich meiner Familie, insbesondere meinen Eltern und Großeltern, dafür danken, dass sie immer an mich geglaubt haben und so zu meiner Motivation beigetragen haben.

Abstract

Since the advent of the internet, several aspects of journalism have been assisted by computers, including data analysis, data collection and organization, and news distribution. This also led to the emergence of Automated Journalism (AJ), which is the automated production of news by automated text generation. The aim of this thesis is to provide news organizations with crucial insights on how the emergence of AJ can change the way how to differentiate themselves from their competition, serving as a basis for their strategic decision making. To accomplish this, four primary areas were investigated in the light of AJ: chances and opportunities, risks and challenges, legal implications, and differentiating characteristics of news organizations. The applied methodology is a comprehensive integrative literature review to establish the theoretical foundation, supplemented by semi-structured interviews with key personnel from news companies in the DACH region. Automated news generation allows news outlets to gain more subscribers and to increase their advertising revenue by drastically increasing their coverage. The increased breadth, specificity, and immediacy of automated news paves the way for almost real-time coverage, even for niche topics. Ideal conditions provided, algorithms are also less error-prone compared to humans. Moreover, AJ makes it possible to publish in multiple languages simultaneously and to increase customer loyalty by personalizing stories. A potential challenge can emerge from company size because small companies tend to struggle with providing enough resources for the introduction of an AJ system. Another challenge is to provide readers with the right balance of transparency and to keep bias to a minimum. The potential for malicious manipulation also poses risks. For the legal aspect, there is no definite consensus on who should be the legal original author of automated news and whether copyright protection is applicable or not. With respect to system transparency, EU countries have to consider that the GDPR requires some degree of explanation for automated decisions. News companies can differentiate themselves to gain a competitive advantage by being innovative and building their own expertise on the topic. They should try to become their own data hosts to provide their readers with unique stories and cut out on external third-party vendors to strengthen their autonomy. This also means to commit to in-house development, provided that enough IT resources are at hand and that they want to be active in the field for the long term. Otherwise, third-party solutions are a good way to get fast results with low initial costs. While currently rule-/template based AJ solutions are dominating real-world use, innovative ML-based solutions hold the potential for future disruptive changes.

Kurzfassung

Seit dem Aufkommen des Internets werden verschiedene Aspekte des Journalismus von Computern unterstützt, darunter die Datenanalyse, die Datenerfassung und -organisation sowie die Nachrichtenverbreitung. Dies führte auch zum Aufkommen von Automated Journalism (AJ), d. h. der automatisierten Produktion von Nachrichten durch automatische Textgenerierung. Ziel dieser Arbeit ist es, als Grundlage für die strategische Entscheidungsfindung von Nachrichtenorganisationen zu dienen, indem wichtige Erkenntnisse darüber vermitteln werden, wie das Aufkommen von AJ die Art und Weise verändern kann, wie sie sich von ihren Konkurrenten abheben können. Um dies zu erreichen, wurden vier Hauptbereiche im Hinblick auf AJ untersucht: Chancen und Möglichkeiten, Risiken und Herausforderungen, rechtliche Implikationen und Differenzierungsmerkmale von Nachrichtenunternehmen. Die angewandte Methodik besteht aus einem umfassenden Integrative Literature Review, um die theoretischen Grundlagen zu schaffen, ergänzt durch Semi-Structured Interviews mit Schlüsselpersonen aus Nachrichtenunternehmen in der DACH-Region. Die automatisierte Nachrichtengenerierung ermöglicht es Nachrichtenunternehmen, mehr Abonnenten zu gewinnen und ihre Werbeeinnahmen zu steigern, durch eine drastische Ausdehnung ihrer Berichterstattung. Die größere Breite, Spezifität und Unmittelbarkeit von automatisierten Nachrichten ebnet den Weg für eine Berichterstattung in nahezu Echtzeit, selbst bei Nischenthemen. Unter idealen Bedingungen sind Algorithmen auch weniger fehleranfällig als Menschen. Darüber hinaus ermöglicht AJ die gleichzeitige Veröffentlichung in mehreren Sprachen und eine stärkere Kundenbindung durch die Personalisierung von Nachrichten. Eine potenzielle Herausforderung kann sich aus der Unternehmensgröße ergeben, denn kleine Unternehmen haben oft Schwierigkeiten, genügend Ressourcen für die Einführung eines AJ-Systems bereitzustellen. Eine weitere Herausforderung besteht darin, den Lesern das richtige Maß an Transparenz zu bieten und Bias zu minimieren. Auch das Potenzial für böswillige Manipulationen birgt Risiken. Was den rechtlichen Aspekt anbelangt, so besteht kein eindeutiger Konsens darüber, wer der rechtliche Urheber automatisierter Nachrichten sein sollte und ob ein Urheberrechtsschutz anwendbar ist oder nicht. In Bezug auf die Systemtransparenz müssen EU-Länder berücksichtigen, dass die Datenschutz-Grundverordnung ein gewisses Maß an Erklärung für automatisierte Entscheidungen verlangt. Nachrichtenunternehmen können sich abheben und einen Wettbewerbsvorteil erlangen, indem sie innovativ sind und ihr eigenes Fachwissen zu diesem Thema aufbauen. Sie sollten versuchen, ihre eigenen Datenhosts zu werden, um ihren Lesern einzigartige Geschichten bieten zu können und auf externe

Drittanbieter verzichten, um ihre Autonomie zu stärken. Das bedeutet auch, dass eine interne Softwareentwicklung bevorzugt werden sollte, sofern genügend IT-Ressourcen zur Verfügung stehen und ein langfristiges Commitment gegeben ist. Andernfalls sind Lösungen von Drittanbietern ein guter Weg für schnelle Ergebnisse bei niedrigen Initialkosten. Während derzeit regelbasierte/templatebasierte AJ-Lösungen in der Praxis vorherrschend sind, bergen innovative ML-basierte Lösungen das Potenzial für künftige disruptive Veränderungen.

Contents

CHAPTER 1

Introduction

Traditionally, the news industry and journalism in general has always been a very involved manual task with a lot of manpower needed. Research and story writing has been done by journalists without computers involved.[1] The first time computers entered some newsrooms was in the end of the 1970s, when they were used to replace traditional typewriters.[2] The advent of computers in journalism can be considered as a major turning point after the invention of printing.[3] During the late 1980s, journalists gradually started to use computers for analyses or to help gathering journalistic knowledge.[4] At the same time, technologies like tele- and radiotext emerged.[5] However, the biggest revolution towards digital journalism came with the advent of the internet in the mid 1990s.[6],[7] From this point on, news organizations started the process of digitalizing the whole news industry, with electronic publication rapidly growing in popularity and the internet becoming the main tool for communication and gathering information.[8] Several aspects of journalism have been assisted by computers over the years, including data analysis, data collection and organization as well as news distribution.[9] By the late 2000s, the ever-advancing trend of digitalization and the emergence of new and innovative technologies, especially Artificial Intelligence (AI), has led to the phenomenon of Automated Journalism

---

[1]See Carlson, "Automated journalism: A posthuman future for digital news?", p. 226.
[2]See Salaverría, "Digital Journalism", p. 3.
[3]See Saad and Issa, "Integration or Replacement: Journalism in the Era of Artificial Intelligence and Robot Journalism", p. 1.
[4]See Caswell and Anderson, "Computational Journalism", p. 2.
[5]See ibid., p. 2.
[6]See Salaverría, "Digital Journalism", p. 3.
[7]See Caswell and Anderson, "Computational Journalism", p. 2.
[8]See Saad and Issa, "Integration or Replacement: Journalism in the Era of Artificial Intelligence and Robot Journalism", p. 3.
[9]See Graefe, *Guide to Automated Journalism*, p. 14.

(AJ).[10],[11] It is the process of automating the production of news along the whole process chain of news generation, from data all the way to publishing the finished story.[12] This means that news organizations are confronted with upcoming technologies that have the potential to re-shape the whole news industry in the long term.[13]

Today's journalists are facing increasing market pressure, with many newsrooms having to cut back on resources due to commercial pressure. Consequently, journalists have to deliver more stories in ever shorter timeframes.[14] Readers add to this pressure even more by expecting prompt news coverage after interesting events.[15] This is one reason why news outlets tend to be cautiously optimistic about the potentials of automated news generation.[16] Beside the automated production of news, there are also several other areas of application for algorithms in the news industry, like automated moderation of comments[17], data mining, detection of trending topics,[18],[19] data visualization[20], or automated fact-checking[21]. The focus of this thesis is the automated production of news by automatically generated text.

AJ can still be considered in its early stages, with a strong growth of scientific publications since 2012.[22] It is an interdisciplinary research field that is currently dominated by journals in the field of journalism and communication. Computer science journals are underrepresented so far.[23] Countries most active in researching AJ are the U.S., Germany, and Spain.[24] A problem not yet illuminated by the state-of-the-art literature is the impact of AJ on the differentiating characteristics of news organizations. The way news organizations currently differentiate themselves from their competitors might be subject to change with the ongoing adoption of AJ. Therefore, it is crucial to understand how news organizations can differentiate themselves in the age of AJ. Intelligent technologies come with great chances and opportunities for news companies as well as with risks and

---

[10]See Caswell and Anderson, "Computational Journalism", p. 2.

[11]See Saad and Issa, "Integration or Replacement: Journalism in the Era of Artificial Intelligence and Robot Journalism", p. 1.

[12]See Carlson, "Automated journalism: A posthuman future for digital news?", p. 226.

[13]See Caswell, "Structured Journalism and the Semantic Units of News", pp. 1, 2.

[14]See Komatsu et al., "AI Should Embody Our Values: Investigating Journalistic Values to Inform AI Technology Design", p. 1.

[15]See Ombelet, Kuczerawy, and Valcke, "Employing Robot Journalists: Legal Implications, Considerations and Recommendations", p. 731.

[16]See Komatsu et al., "AI Should Embody Our Values: Investigating Journalistic Values to Inform AI Technology Design", p. 1.

[17]See Miroshnichenko, "AI to bypass creativity. Will robots replace journalists? (The answer is "yes")", p. 4.

[18]See ibid., p. 4.

[19]See S. Wu, Tandoc Jr, and Salmon, "A Field Analysis of Journalism in the Automation Age: Understanding Journalistic Transformations and Struggles Through Structure and Agency", p. 433.

[20]See ibid., p. 433.

[21]See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", pp. 45, 46.

[22]See Xu and Lan, "A Scientometric Review of Automated Journalism: Analysis and Visualization", pp. 2, 6.

[23]See ibid., pp. 1, 4.

[24]See ibid., p. 1.

challenges that have to be taken into account. Algorithmic transparency, the risk of bias and ethical considerations are, among others, important topics that have to be considered on this matter. Legal aspects concerning fundamental rights, authorship, and liability are also essential subjects of research when considering the real-world adoption of AJ in newsrooms. The aim of this thesis is to give news organizations crucial insights on how the emergence of AJ can change the way how to differentiate themselves from their competition. Such knowledge is important for the news industry because it can serve as a basis for strategic decision making for news organizations regarding the implementation of new technologies related to AJ. Well-considered strategic decisions of this sort can determine their future competitive edge. The following research questions were used to explore the problem:

1. What are the chances and opportunities of AJ for news organizations?

2. What are the risks and challenges of AJ for news organizations?

3. What are the legal implications of AJ with respect to authorship, liability, and fundamental rights, with focus on the European Union (EU)?

4. In which way do these impacts of AJ influence the differentiating characteristics of news organizations?

This thesis aims to answer the above-mentioned research questions step by step by applying a comprehensive integrative literature review, supplemented by semi-structured interviews. The overall structure of the thesis is based on the research questions. Chapter 2 describes the methodological approach in detail. Chapters 3 to 6 establish foundational knowledge for AJ and its related technologies as well as its current state and readers' perception. Chapters 7 to 10 are directly related to the research questions, covering the changes and opportunities of AJ, its risks and challenges, legal implications, and differentiating characteristics of news organizations.

CHAPTER 2

Methodological Approach

The methodology follows a twofold approach. First, the theoretical foundation of the topic is established through a comprehensive integrative literature review, based on related research that has already been carried out in the area of AJ. This is a major part of the thesis. Second, qualitative research in the form of semi-structured interviews has been conducted. The aim is to get an understanding on how news organizations in the DACH region (Germany, Austria, and Switzerland) are currently making use of advanced technological systems, what they may find most challenging, how they are currently differentiating themselves from their competitors, and how they think about future developments. In case it has not been possible to conduct semi-structured interviews in person or via video or phone calls, written open-ended questions have been sent via e-mail as an alternative medium for information retrieval.

2.1 Integrative Literature Review

According to Snyder, the integrative literature review, compared to systematic or semi-systematic reviews, allows for more flexibility in the collection of relevant data. This is why integrative reviews are best suited when the main goal is to combine different perspectives to derive new insights for a particular research question, which means that covering all papers ever published on a topic is not the main purpose. This type of review should result in the advancement of knowledge by answering a particular research question rather than just an overview of a research area.[1] In this thesis the integrative literature review has been used because it is the best fit for the problem that is supposed to be investigated. Different perspectives on the field of AJ have to be combined, which is only possible with the flexibility that an integrative review approach allows. To guarantee depth, rigor and transparency, a four-phase process proposed by Snyder has been used[2].

---

[1]See Snyder, "Literature review as a research methodology: An overview and guidelines".
[2]See ibid.

5

Its phases are design, conduct, analysis, and structuring and writing the review[3]. All relevant decisions related to the literature review have been documented to allow for an understanding of how the literature has been identified, analyzed, and synthesized.

In a first step, search terms, databases to be used, and appropriate inclusion and exclusion criteria have been specified. The list of search terms consists of "automated journalism" and related terms like "robot journalism", "computational journalism", and "algorithmic journalism". Other terms investigated are "natural language generation journalism", "automated text generation journalism", "algorithmic bias", "algorithmic accountability", "journalism ethics", and "algorithmic authorship". Aspects like challenges related to software developing and the potential dependency on software suppliers have also been considered. A clear exclusion criterion for literature is when only special prototype applications were developed by the authors. Very technical papers from the computer science domain also have not been considered. There are five scientific libraries that have been used for the research: dblp, IEEE Xplore, ACM DL, Scopus, and Google Scholar. For the legal research, the Austrian legal information system "RIS" as well as databases for European law like "EUR-Lex" and the "European e-Justice Portal" have primarily been used.

In a second step, a pilot has been carried out to guarantee for an appropriate review process. The literature selection has been conducted in stages. First, the titles have been read and an initial selection has been made. After that, all abstracts have been read to further refine the selection. For the final selection, the full text has been taken into account. This way it is guaranteed that the literature has been critically analyzed and the whole process is transparent and well documented. After considering the titles, 203 papers ended up in the selection. With abstracts considered, 129 paper remained, while 102 formed the final selection for this thesis.

2.2 Semi-Structured Interviews

In this thesis, Semi-Structured Interviews (SSI) are used to gain more practical insights from actual news organizations on how they cope with the area of AJ at the moment and about their differentiating characteristics over time. The purpose of these interviews is to add more depth by investigating the news sector in the DACH region, in addition to the primary method, the integrative literature review. It is not the goal to establish a representative coverage of all DACH news companies. Instead, the semi-structured interviews are used to incorporate valuable first-hand knowledge from the industry in order to supplement the literature review by adding more depth on certain questions. According to Adams, the SSI is an approach that falls between highly structured surveys that mostly consist of closed-ended questions on the one hand and highly fluid focus groups on the other hand[4]. This means that SSIs are a mixture of closed- and open-ended questions. They are well suited to add depth in mixed methods research along with other approaches.

---

[3]See Snyder, "Literature review as a research methodology: An overview and guidelines".
[4]See Adams, "Conducting Semi-Structured Interviews".

SSIs allow to study people's opinions and perceptions on complex topics.[5] Kallio et al. state that the first step in conducting a SSI is "to gain a comprehensive and adequate understanding of the subject"[6], which has been achieved with the integrative literature review. For the SSIs, a three-step process has been followed: interview setup/design, conducting the interviews, and post-analysis. In the final step of the design phase an interview guideline has been created, which means planning the topics to be covered, coming up with questions to be addressed, and to decide on a tentative order.[7] The interview questions are split into main and follow-up questions. While the main themes are there to cover the main content and to provide a logical order of the topics, follow-up questions serve the purpose of guiding the conversation by better understanding the main theme and to get more accurate information, while also being beneficial for the fluency of the conversation.[8] The questions used for the interviews are the following:

Interview Guideline [EN]

1. Does [company name] currently use intelligent technologies for the automation of workflows, particularly automated news generation? If so, which ones?

    1.1 *(1.1-1.4: If technology for automation is in use)*
    Is the software that is in use developed by your company itself or is it licensed from a third-party supplier?

    1.1.1 *(If licensed)*
    Was it an option to start developing one's own software? Why (not)?

    1.1.2 *(If self-developed)*
    What were the reasons for an in-house development? What benefits and challenges emerged from this?

    1.2 From your current perspective, what is your company's biggest challenge when it comes to making one's own software solutions?

    1.3 What are the biggest benefits that come along with the software solution deployed in your company?

    1.4 What are the biggest challenges or downsides that come along with the software solution deployed in your company?

2. *(If no automation tools adopted yet)*
    Has the introduction of software automation tools ever been considered in your company so far? If so, which ones?

---

[5]See ibid.

[6]Kallio et al., "Systematic methodological review: developing a framework for a qualitative semi-structured interview guide".

[7]See Adams, "Conducting Semi-Structured Interviews".

[8]See Kallio et al., "Systematic methodological review: developing a framework for a qualitative semi-structured interview guide".

   2.1 Can you think of workflows in a newsroom that would be well suited to be supported or taken over by intelligent technologies?

3. *(If no automated news generation adopted yet)*
In your opinion, what are the key reasons that kept [company name] from adopting automated news generation so far?

4. In your opinion, how does [company name] currently differentiate itself from competitors?

   4.1 With an extensive adoption of automated journalism in the future, how might this current key differentiating characteristics be subject to change?

Interview Guideline [DE]

1. Nutzt [Unternehmen] aktuell intelligente Technologien für die Automation von Workflows, insbesondere zur automatischen Generierung von News? Wenn ja, welche?

   1.1 *(1.1-1.4: Wenn Technologie zur Automatisierung in Verwendung)*
Ist die verwendete Software eine Eigenentwicklung Ihres Unternehmens oder wird sie von einem Drittanbieter lizensiert?

      1.1.1 *(Wenn lizensiert)*
War es eine Option, die Software selbst zu entwickeln? Warum (nicht)?

      1.1.2 *(Wenn Eigenentwicklung)*
Was waren die Gründe für eine Eigenentwicklung? Welche Vorteile und Herausforderungen haben sich dadurch ergeben?

   1.2 Was ist aus heutiger Sicht die größte Herausforderung für Ihr Unternehmen, wenn es um die Entwicklung eigener Softwarelösungen geht?

   1.3 Was sind die größten Vorteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?

   1.4 Was sind die größten Herausforderungen oder Nachteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?

2. *(Wenn bisher noch keine Automatisierungstools in Verwendung)*
Wurde die Einführung von softwarebasierten Automatisierungstools in Ihrem Unternehmen jemals in Betracht gezogen? Wenn ja, welche?

   2.1 Fallen Ihnen Workflows in der Redaktion ein, die sich eignen würden, von intelligenten Technologien unterstützt oder ersetzt zu werden?

3. *(Wenn bisher noch keine automatisierte Generierung von News in Verwendung)*
Was sind Ihrer Meinung nach die Hauptgründe, die [Unternehmen] bisher davon abgehalten haben, die automatisierte Generierung von News einzuführen?

4. Wie differenziert sich [Unternehmen] Ihrer Meinung nach aktuell von Konkurrenten?

    4.1 Mit einer zukünftig umfassenden Einführung von Automated Journalism, wie könnten sich diese aktuellen Hauptdifferenzierungsmerkmale künftig verändern?

The interviews have been held in German because all the interviewed companies are German-speaking. This means that the answers used in the thesis have been translated to English afterwards. In the last step of developing our interview guideline the questions were arranged according to their importance, to ensure that the most important ones get covered within time[9]. One peculiarity of SSIs is that the interview guide is just a guideline and the interview itself is loose and flexible. This means that new and relevant topics can come up during the interview that can be further discussed. If necessary, questions have been re-ordered on the fly during the interview.[10],[11] Once the interview guide was finished, an internal pilot test with fellow students has been conducted.[12] In case it has not been possible to conduct semi-structured interviews in person or via video or phone calls, the alternative plan for information retrieval was to send out written questions via e-mail to the same target group. These questions were based on the interview guideline that was developed for the SSI.

Table 2.1: List of participants

Code	Position
A	Data Journalist, Head of Development
B	Data Journalist
C	Department Head
D	Head of Development
E	Data & Automation Specialist
F	Editor in Chief, Digital Innovation
G	Team Lead - Automated Journalism
H	Chief Product Officer

The target group are editors and journalists from different German-speaking news organizations who are in touch with the technologies used in their company. Respondents can be from different hierarchical levels, with a focus on the management level. All interviewees should have an overview of the software solutions used in their company, which are used in everyday editorial work. An ideal respondent would be key personnel who is involved in the management or development of software solutions for the automation of workflows in the company. However, most of the interview questions can be answered

---

[9]See Adams, "Conducting Semi-Structured Interviews".
[10]See ibid.
[11]See Kallio et al., "Systematic methodological review: developing a framework for a qualitative semi-structured interview guide".
[12]See ibid.

without in-dept knowledge about a company's IT solutions. In the end, between April and May 2022, 8 interviews were conducted, see Table 2.1. Each interviewee is working at a different company, meaning there is no overlap in companies. According to McCracken, having 8 respondents from the desired target group can be considered ideal in most cases[13]. The criteria for an ideal respondent defined above are met by 7 out of 8 of our participants. Furthermore, the interviewed news organizations are quite evenly distributed across the DACH region, with 3 being from Germany, 3 from Switzerland, and 2 from Austria. All except one of the interviews were held via Zoom or telephone and took between 15 to 20 minutes. The audio has been recorded to create a transcript afterwards. Alphabetical codes have been assigned to each SSI and sensible information was anonymized so that no conclusions can be drawn about the identity of the persons or companies involved. All transcripts can be found in the appendix, see chapter *Appendix - Interview Transcripts*. The questions of interview H have been answered in writing. In the final step, all the interview results have been analyzed by identifying relevant information and consolidating themes found in multiple answers[14]. The final findings have been cited in the respective parts of this thesis. By adhering to this proposed guideline for the process of conducting a SSI, rigor and trustworthiness are ensured[15].

---

[13]See McCracken, *The Long Interview*, p. 17.
[14]See Adams, "Conducting Semi-Structured Interviews".
[15]See Kallio et al., "Systematic methodological review: developing a framework for a qualitative semi-structured interview guide".

CHAPTER 3

Automated Journalism

Over the years, several terms for the automated generation of news emerged in the scientific literature. The most popular ones are "automated journalism"[1,2,3], "computational journalism"[4,5,6,7], "algorithmic journalism"[8,9], and "robot journalism"[10,11,12,13,14]. All of these terms refer to the same concept, which is why they can be used interchangeably. However, in this thesis the term Automated Journalism (AJ) will be used because it probably reflects the underlying nature of automation in the news generation process best, without the risk of inducing misleading expectations like it could be the case with the term "robot journalism". According to Carlson, the term automated journalism "[...] denotes algorithmic processes that convert data into narrative news texts with limited to no human intervention beyond the initial programming choices."[15] This definition is

---

[1] See Carlson, "The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority".

[2] See Graefe, *Guide to Automated Journalism*.

[3] See Caswell and Dörr, "Automated Journalism 2.0: Event-driven narratives: From simple descriptions to real stories".

[4] See Anderson, "Notes Towards an Analysis of Computational Journalism".

[5] See Anderson, "Towards a sociology of computational and algorithmic journalism".

[6] See Flew et al., "The promise of computational journalism".

[7] See Coddington, "Clarifying Journalism's Quantitative Turn".

[8] See Anderson, "Towards a sociology of computational and algorithmic journalism".

[9] See Dörr, "Mapping the field of Algorithmic Journalism".

[10] See Clerwall, "Enter the Robot Journalist: Users' perceptions of automated content".

[11] See Dalen, "The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists".

[12] See Latar, "The Robot Journalist in the Age of Social Physics: The End of Human Journalism?"

[13] See Montal and Reich, "I, Robot. You, Journalist. Who is the Author?: Authorship, bylines and full disclosure in automated journalism".

[14] See Thurman, Dörr, and Kunert, "When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism's capabilities and consequences".

[15] Carlson, "The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority", p. 416.

11

particularly well suited for the domain because it precisely captures the essence of the subject, the automated production of news. It encompasses the whole process chain of news production, starting with the data collection and analysis, followed by the text generation itself, and ending in the news publication.[16] The term *automated news* will be used to refer to the news texts created by AJ.

Algorithms can be described as a set of sequential instructions that are executed by a computer to achieve a desired outcome.[17] In order to produce automated news, algorithms have to perform several steps in the automation process. First, the system needs input in the form of raw data that comprises all the necessary information. Pre-processing might be needed to get clean and structured data for further analysis. Next, relevant information is extracted and classified by importance by applying statistical heuristics. Now a text structure can be planned by arranging the newsworthy content or a pre-defined template gets filled accordingly. Following from this semantic structure, a human readable text, in the form of meaningful grammatically correct sentences, has to be created.[18][19][20] The final step is the publishing of the finished news text. This final publication can also be automated by including the newsroom's content management system in the process.[21]

3.1 Explanatory Models

Graefe came up with a model that describes the different steps of automated news generation in detail, see figure 3.1. After the crucial fundamental stage of data collection, interesting events have to be identified. Both statistical analysis and pre-defined domain specific rules are used for this identification. The former reveals interesting trends, correlations and outliers, while the latter are adapted for the specific problem at hand. The development of such pre-defined rules poses an interdisciplinary challenge that involves both computer scientists and journalists. Next, the identified insights have to be prioritized regarding their newsworthiness. Again, a collaboration with domain experts is needed to come up with rules the algorithm can use to rank the insights. The final text generation further requires computer linguists. They need to develop a rule-based system that generates grammatically correct sentences, in accordance to the publisher's official style guide. Publishing the story is the final step and can be fully automated or require editorial review first.[22]

On the other hand, Dörr proposes a more generalistic model on how Natural Language Generation (NLG) can be used for AJ, the Input-Throughput-Output (I-T-O) model,

---

[16]See Graefe, *Guide to Automated Journalism*, p. 14.
[17]See Linden, "Algorithms for journalism: The future of news work", p. 62.
[18]See Haim and Graefe, "Automated News: Better than expected?", pp. 102, 103.
[19]See Kaa and E. J. Krahmer, "Journalist versus news consumer: The perceived credibility of machine written news", p. 2.
[20]See Montal and Reich, "I, Robot. You, Journalist. Who is the Author?: Authorship, bylines and full disclosure in automated journalism", pp. 3, 4.
[21]See Haim and Graefe, "Automated News: Better than expected?", pp. 102, 103.
[22]See Graefe, *Guide to Automated Journalism*, pp. 17–19.
[23]Based on Graefe, *Guide to Automated Journalism*, p. 18.

Figure 3.1: Graefe's model on how algorithms generate news[23]

see figure 3.2. It describes the process of automated news generation on a more abstract level.[24] Like the model's name already suggests, input in the form of data gets transformed into natural language text. The throughput phase consists of two major steps, (1) text planning and (2) linguistic realization. Text planning encompasses the identification and priorization of interesting events, resulting in a document structure. During the linguistic realization this document structure gets transformed into natural language by the NLG algorithm. For instance, the algorithm has to decide for sentence structures and make syntactic choices concerning the words to be used for the particular domain.[25][26] There are also feedback loops involved in the process, representing the human influence or intervention until the outcome meets the desired result, thus allowing for a dynamic flow of information.[27][28][29]

However, both Graefe's and Dörr's models only account for template-based automated

---

[24]See Dörr, "Mapping the field of Algorithmic Journalism", pp. 703–705.
[25]See ibid., pp. 703–705.
[26]See Dalgali and Crowston, "Algorithmic Journalism and Its Impacts on Work", p. 2.
[27]See Dörr, "Mapping the field of Algorithmic Journalism", pp. 703–705.
[28]See Dalgali and Crowston, "Algorithmic Journalism and Its Impacts on Work", p. 2.
[29]See Habel, *Roboterjournalismus*, p. 40.

text generation. This only allows for more simple and descriptive news in limited domains, while human journalists are able to write much more complex news articles in all possible domains.[30] Caswell and Dörr addressed this problem by proposing a model that makes use of semantic features, based on a "story database", to produce event-driven narratives.[31,32] "Journalists enter events and narratives into this database, which uses the semantics, or meanings, of journalistic events to categorize news stories in the form of structured data."[33] This event-driven approach, in theory, allows for more complex automated news compared to merely template-based approaches.[34]
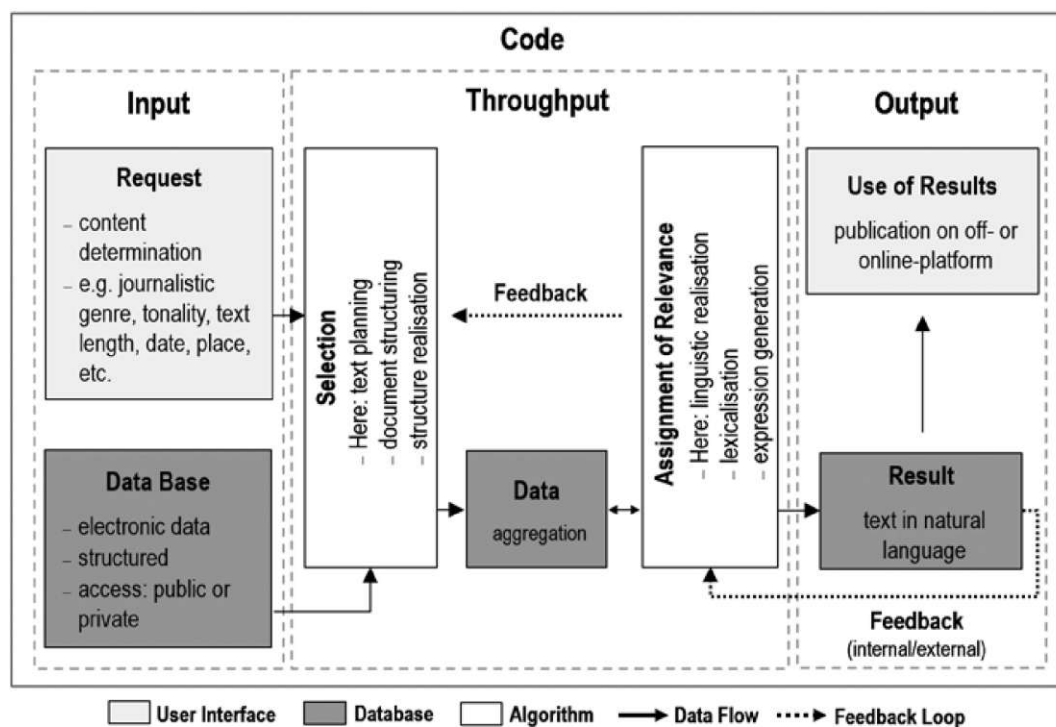
Figure 3.2: Dörr's Input-Throughput-Output (I-T-O) model[35]

---

[30]See Dalgali and Crowston, "Algorithmic Journalism and Its Impacts on Work", p. 2.

[31]See Caswell and Dörr, "Automated Journalism 2.0: Event-driven narratives: From simple descriptions to real stories".

[32]See Dalgali and Crowston, "Algorithmic Journalism and Its Impacts on Work", p. 2.

[33]Ibid., p. 2.

[34]See Caswell and Dörr, "Automated Journalism 2.0: Event-driven narratives: From simple descriptions to real stories", p. 16.

[35]Dörr, "Mapping the field of Algorithmic Journalism", p. 704.

Related Technologies

For the technical realization, Automated Journalism utilizes subfields of Artificial Intelligence. The term AI exists since the 1950s and is nowadays associated with various meanings.[1] A general definition of AI is that it pursues the goal to emulate intelligent human behavior through computational processes.[2,3,4] In the context of AJ, Artificial Intelligence can be described as "[...] a system's ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation."[5] It is worth mentioning that AI tools used in newsrooms today are mostly general purpose, originating from other disciplines, like, for example, engineering, statistics, or computer science.[6] AI has several subfields, with Machine Learning (ML) and Natural Language Processing (NLP) being most important for AJ.[7,8] ML is defined as "[...] any methodology and set of techniques that finds novel patterns and knowledge in data, and generates models (eg profiles) that can be used for effective predictions about the data."[9] Besides the generation of automated news, AI tools are also used for things like breaking news tracking, enhanced searches and information linking, automated generation of graphics from data, fake news discovery, or automated grammar and spelling checks.[10] AI algorithms are also used to derive useful information from large

---

[1]See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, p. 2.
[2]See Vázquez-herrero et al., *Journalistic Metamorphosis: Media Transformation in the Digital Age*, p. 20.
[3]See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, p. 2.
[4]See Saad and Issa, "Integration or Replacement: Journalism in the Era of Artificial Intelligence and Robot Journalism", p. 6.
[5]Kaplan and Haenlein, "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence", p. 17.
[6]See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, p. 2.
[7]See Broussard et al., "Artificial Intelligence and Journalism", p. 5.
[8]See Biswal and Gouda, "Artificial Intelligence in Journalism: A Boon or Bane?", p. 156.
[9]Otterlo, "A machine learning view on profiling", p. 46.
[10]See Saad and Issa, "Integration or Replacement: Journalism in the Era of Artificial Intelligence and Robot Journalism", p. 10.

data silos through data analytics, also called data mining.[11],[12]

4.1 Natural Language Generation

One of the most important drivers of AJ is Natural Language Generation (NLG)[13],[14],[15],[16], which is a subfield of Natural Language Processing (NLP)[17],[18]. NLG is a technology that allows for the automatic transformation of structured data (input) into natural, human readable, coherent text (output).[19],[20],[21],[22],[23],[24] The target language can be any human language of choice that is supported by the NLG system.[25],[26] In the 1950s, NLG was first used for machine translation.[27],[28] It is only since the 1980s that NLG became an independent research subfield.[29] Since then, it continued to grow in popularity and is being utilized by different industries, including the news media sector.[30],[31]. Besides for AJ, it is also commonly used for written weather forecasts as well as in the sport, finance, medical, and engineering domain.[32] According to Gatt and Krahmer, the process of Natural Language Generation can be summarized in six consecutive steps: (1) "content determination", (2) "text structuring", (3) "sentence aggregation", (4) "lexicalization",

---

[11]See Saad and Issa, "Integration or Replacement: Journalism in the Era of Artificial Intelligence and Robot Journalism", p. 7.

[12]See Latar, "The Robot Journalist in the Age of Social Physics: The End of Human Journalism?", p. 68.

[13]See Dalgali and Crowston, "Algorithmic Journalism and Its Impacts on Work", p. 1.

[14]See Elikci, "Integration of Artificial Intelligence Into Austrian Journalism: A Potential Threat to Austrian Journalism?", p. 184.

[15]See Kotenidis and Veglis, "Algorithmic Journalism—Current Applications and Future Perspectives", p. 247.

[16]See Dörr, "Mapping the field of Algorithmic Journalism", pp. 700, 702.

[17]See ibid., pp. 700, 702.

[18]See Dalgali and Crowston, "Algorithmic Journalism and Its Impacts on Work", p. 1.

[19]See Utasch and Loebbecke, "Intelligent technologies shaping business models for journalistic content provision: A concept matrix", p. 2.

[20]See Gatt and E. Krahmer, "Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation", p. 68.

[21]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 11.

[22]See Dalgali and Crowston, "Algorithmic Journalism and Its Impacts on Work", p. 1.

[23]See Elikci, "Integration of Artificial Intelligence Into Austrian Journalism: A Potential Threat to Austrian Journalism?", p. 184.

[24]See Caswell and Dörr, "Automated Journalism 2.0: Event-driven narratives: From simple descriptions to real stories", p. 1.

[25]See Gatt and E. Krahmer, "Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation", p. 68.

[26]See Utasch and Loebbecke, "Intelligent technologies shaping business models for journalistic content provision: A concept matrix", p. 2.

[27]See Dörr, "Mapping the field of Algorithmic Journalism", p. 703.

[28]See Kotenidis and Veglis, "Algorithmic Journalism—Current Applications and Future Perspectives", p. 247.

[29]See Dörr, "Mapping the field of Algorithmic Journalism", p. 703.

[30]See ibid., p. 703.

[31]See Kotenidis and Veglis, "Algorithmic Journalism—Current Applications and Future Perspectives", p. 247.

[32]See Dörr, "Mapping the field of Algorithmic Journalism", p. 703.

(5) "referring expression generation", and (6) "linguistic realization".[33] This means that relevant content has to be put into the right order within a cohesive sentence structure. The lexicalization step is the most important one because here the most suitable words and phrases have to be picked. To avoid repeated use of the same words, alternative terms and phrases, referring to the same entities, are determined in step five. In the final linguistic realization everything is put together for the finished output text.[34],[35]

NLG systems can be divided into two main categories, *rule-based* and *machine learning based*.[36] Rule-based systems produce text based on a specific set of rules, defined by humans. These rules have to be specific to the respective domain, thus cannot be transferred to other use cases later on. What makes this approach very time-consuming is that the rules have to cover every possible outcome the data could induce.[37] Rule-based NLG systems are often referred to as template-based systems because in the end these rules form a kind of skeleton text that encompasses variables that have to be filled by the input data.[38] But these templates must not only be simple "gap texts", but can also span highly sophisticated hierarchies of complex logic, determining the order and text fragments to be used, all depending on the actual input data. Unfortunately, this also means that every story template is dependent on a specific data model the input data set has to comply with.[39] Another downside is that complex templates can quickly become hard and time consuming to manage.[40]

Machine learning based NLG systems on the other hand rely on training data as input, instead of rules pre-defined by humans. This approach is especially well suited for domains where many already written articles exist.[41] The NLG system builds a model on the basis of the input data set and the finished training articles representing the desired outcome. This kind of ML is called *supervised machine learning*.[42],[43] However, the requirement for huge amounts of training data can limit its application to more simple domains, where a large number of articles is already available. Also, rules that are not prevalent in the training data cannot be learned, which means the training data has to be selected

---

[33] See Gatt and E. Krahmer, "Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation", pp. 70, 71.

[34] See ibid., pp. 70–79.

[35] See Cremaschi et al., "Supporting journalism by combining neural language generation and knowledge graphs", pp. 4, 5.

[36] See Sirén-Heikel et al., "Unboxing news automation: Exploring imagined affordances of automation in news journalism", p. 49.

[37] See ibid., p. 49.

[38] See Leppänen, Tuulonen, and Sirén-Heikel, "Automated journalism as a source of and a diagnostic device for bias in reporting", p. 42.

[39] See Caswell and Dörr, "Automated Journalism 2.0: Event-driven narratives: From simple descriptions to real stories", p. 3.

[40] See ibid., p. 4.

[41] See Sirén-Heikel et al., "Unboxing news automation: Exploring imagined affordances of automation in news journalism", p. 49.

[42] See Leppänen, Tuulonen, and Sirén-Heikel, "Automated journalism as a source of and a diagnostic device for bias in reporting", p. 43.

[43] See Boumans and Trilling, "Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars", pp. 13, 14.

carefully.[44] Systems based on unsupervised machine learning, which is not dependent on training data in the same way, are very rare and not yet used for AJ applications because of great difficulties that must yet be overcome.[45,46]

Compared to rule-based NLG systems, whose applications are already well researched and proven, ML-based systems offer a much higher potential for future disruptive changes.[47] Nevertheless, ML-based systems are still accompanied by major challenges. The biggest problems are that 100 percent factual accuracy is not yet possible and these systems are also prone to data bias caused by biased training data. A prominent example for the lack of factual accuracy is the ML-based hockey news bot "Scoopmatic", deployed by the Finnish news agency SST. While this bot shined in handling the Finnish language, it only achieved a factual correctness of 85 percent for its generated content.[48] Another problematic phenomena with ML-based NLG systems is that they sometimes tend to "hallucinate" in a way that they come up with informations that are fictitious.[49] Machine learning also brought up the "black box" problematic for the first time, which means that the output is neither *ex ante* predictable nor *ex post* explainable.[50] For these reasons, rule-based systems are mostly without alternative for journalistic applications at present, since they produce output that is ex post explainable.[51,52] This is important because journalism has very high requirements for transparency, quality assurance, and factual correctness.[53]

---

[44]See Sirén-Heikel et al., "Unboxing news automation: Exploring imagined affordances of automation in news journalism", p. 49.

[45]See Leppänen, Tuulonen, and Sirén-Heikel, "Automated journalism as a source of and a diagnostic device for bias in reporting", p. 43.

[46]See Boumans and Trilling, "Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars", pp. 15, 16.

[47]See Sirén-Heikel et al., "Unboxing news automation: Exploring imagined affordances of automation in news journalism", p. 61.

[48]See Krawarik et al., *AI.AT.Media. AI and the Austrian Media Sector: Mapping the Landscape, Setting a Course*, p. 95.

[49]See Schell, *Whitepaper. Journalistische Textautomatisierung. Status, Potenziale, Limitationen*, p. 7.

[50]See Habel, *Roboterjournalismus*, pp. 44, 45, 63.

[51]See Schell, *Whitepaper. Journalistische Textautomatisierung. Status, Potenziale, Limitationen*, p. 7.

[52]See Habel, *Roboterjournalismus*, pp. 44, 45, 63.

[53]See Schell, *Whitepaper. Journalistische Textautomatisierung. Status, Potenziale, Limitationen*, p. 7.

CHAPTER 5

Current State

At the moment, automated news generation is still small-scale and far from mainstream, with only a few news organizations being very active in this domain.[1],[2],[3] Some media outlets already made good progress in the area of AJ[4], while others are still in an experimental phase.[5] When comparing Austria and Germany in terms of automated news generation usage, news organizations in Austria are hardly making use of this technology so far.[6] Wu et al. described the current state of automation in newsrooms on a more generalistic level, including use cases besides solely automated text generation, structured by five stages of news production: gathering, selection, writing, editing, and distribution.[7] For news gathering, algorithms are used for information retrieval, to identify trending topics and to pull relevant videos and images to be used in stories.[8] During news selection, software solutions can help in judging the credibility of news and how newsworthy certain events are, for example, by analyzing their appearance on social media.[9] Besides the actual automated creation of news texts, newsrooms also commonly deploy algorithms for simple tasks like data visualization, translation or transcription.[10] Automated spell

---

[1]See Linden, "Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?", pp. 132, 135.

[2]See Saad and Issa, "Integration or Replacement: Journalism in the Era of Artificial Intelligence and Robot Journalism", p. 7.

[3]See Sirén-Heikel et al., "Unboxing news automation: Exploring imagined affordances of automation in news journalism", p. 63.

[4]See Saad and Issa, "Integration or Replacement: Journalism in the Era of Artificial Intelligence and Robot Journalism", p. 7.

[5]See Zydorek, "Automatisierte Contentproduktion bei Nachrichten – Automated Journalism", p. 108.

[6]See Elikci, "Integration of Artificial Intelligence Into Austrian Journalism: A Potential Threat to Austrian Journalism?", p. 187.

[7]See S. Wu, Tandoc Jr, and Salmon, "Journalism Reconfigured: Assessing human–machine relations and the autonomous power of automation in news production", p. 8.

[8]See ibid., pp. 8, 9.

[9]See ibid., pp. 9, 10.

[10]See ibid., pp. 10, 11.

checks are an example for the news editing phase, although human verification in still important. Finally, for the news publishing it is very common nowadays to automate the online publishing of stories and to make use of auto-scheduling.[11] These examples of automation in different stages of news production show that machines gained substantial influence in this process. Especially in the news gathering step, by identifying trending topics and scraping data, algorithms can set direction on what stories journalists will write. Setting the parameters for this algorithms carefully is therefore key to achieve the desired outcome.[12]

5.1   Practical Applications

Popular domains for automated text generation are those that can easily be represented in a template and where structured data in the form of tables, rankings, etc. is available.[13],[14] Also, recurring topics of cyclic nature[15],[16],[17], with limited internal variability, are particularly suited for AJ[18],[19]. Automation allows for extremely fast publication of news, making it especially useful for topics where speed matters.[20] Current rule-based NLG solutions are best for describing what happened at specific predictable events, unlike complex storytelling.[21],[22],[23] Only few of this solutions are even able to make comparisons between new and historical data, thus being quite limited in their application.[24] For this reasons, at the moment, the most popular domains for AJ are sports, finance, real

---

[11]See S. Wu, Tandoc Jr, and Salmon, "Journalism Reconfigured: Assessing human–machine relations and the autonomous power of automation in news production", pp. 12, 13.

[12]See ibid., pp. 14, 15.

[13]See ibid., p. 11.

[14]See interview transcript C, question 2.1.

[15]See S. Wu, Tandoc Jr, and Salmon, "Journalism Reconfigured: Assessing human–machine relations and the autonomous power of automation in news production", p. 11.

[16]See Luengo and Herrera-Damas, *News Media Innovation Reconsidered*, p. 125.

[17]See interview transcript D, question 1.4.

[18]See Sirén-Heikel et al., "Unboxing news automation: Exploring imagined affordances of automation in news journalism", p. 50.

[19]See interview transcript D, question 1.4.

[20]See Linden, "Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?", p. 131.

[21]See Saad and Issa, "Integration or Replacement: Journalism in the Era of Artificial Intelligence and Robot Journalism", p. 11.

[22]See Sirén-Heikel et al., "Unboxing news automation: Exploring imagined affordances of automation in news journalism", p. 50.

[23]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 11.

[24]See Fanta, "Putting Europe's robots on the map: automated journalism in news agencies", p. 10.

estate, election results, opinion polls, weather, and traffic news.[25][26][27][28][29][30][31] The field automated news generation first emerged from is weather forecasting, where textual forecasts are generated by choosing the right pre-defined phrases, depending on how high certain metrics like temperature or wind speed of a forecasting model are.[32]

One of the first well known examples for the application of automated text generation is the *Los Angeles Times's* project on crime reporting since 2007. The corresponding section of the newspaper is called "The Homicide Report".[33][34] This project was aimed to mitigate the bias caused by covering only about 10 percent of all annual homicides in L.A. County. Thus, the goal for the algorithm was to cover not only the most outstanding and newsworthy homicides, but every singe one occurring, hence resulting in a representative presentation of the topic.[35] The system relies on publicly available structured data and "[...] includes information such as the date, location, time, race or ethnicity, age, jurisdiction, and neighborhood of all homicides in the area."[36] Furthermore, all the data is shown on a county map and allows to filter by the aforementioned datapoints, see figure 5.1. This map also provides relevant insights for the real estate market in the area.[37] However, the short reports generated by the algorithm are not replacing human journalists on this topic. Just like before the introduction of the NLG system, journalists are still covering particularly newsworthy stories by enhancing the short news produced by the software with interesting background information about the victim.[38][39]

The Los Angeles Times is also using a simple but effective AI solution called "Quakebot", which is still in use since 2014.[41] It automatically reports on the location, time, and

[25] See ibid., p. 10.

[26] See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 11.

[27] See Saad and Issa, "Integration or Replacement: Journalism in the Era of Artificial Intelligence and Robot Journalism", pp. 7, 11.

[28] See S. Wu, Tandoc Jr, and Salmon, "Journalism Reconfigured: Assessing human–machine relations and the autonomous power of automation in news production", p. 11.

[29] See Sirén-Heikel et al., "Unboxing news automation: Exploring imagined affordances of automation in news journalism", p. 50.

[30] See Elikci, "Integration of Artificial Intelligence Into Austrian Journalism: A Potential Threat to Austrian Journalism?", p. 184.

[31] See Luengo and Herrera-Damas, *News Media Innovation Reconsidered*, p. 125.

[32] See Graefe, *Guide to Automated Journalism*, p. 19.

[33] See ibid., p. 21.

[34] See Miroshnichenko, "AI to bypass creativity. Will robots replace journalists? (The answer is "yes")", p. 5.

[35] See Graefe, *Guide to Automated Journalism*, p. 21.

[36] Ibid., p. 21.

[37] See Miroshnichenko, "AI to bypass creativity. Will robots replace journalists? (The answer is "yes")", p. 5.

[38] See Graefe, *Guide to Automated Journalism*, p. 22.

[39] See Miroshnichenko, "AI to bypass creativity. Will robots replace journalists? (The answer is "yes")", p. 5.

[40] Screenshot from https://homicide.latimes.com/year/2021, accessed on 05/03/2022.

[41] See Kunert, "Journalists, meet your new colleague algorithm: The impact of automation on content distribution and content creation in the newsroom", p. 140.
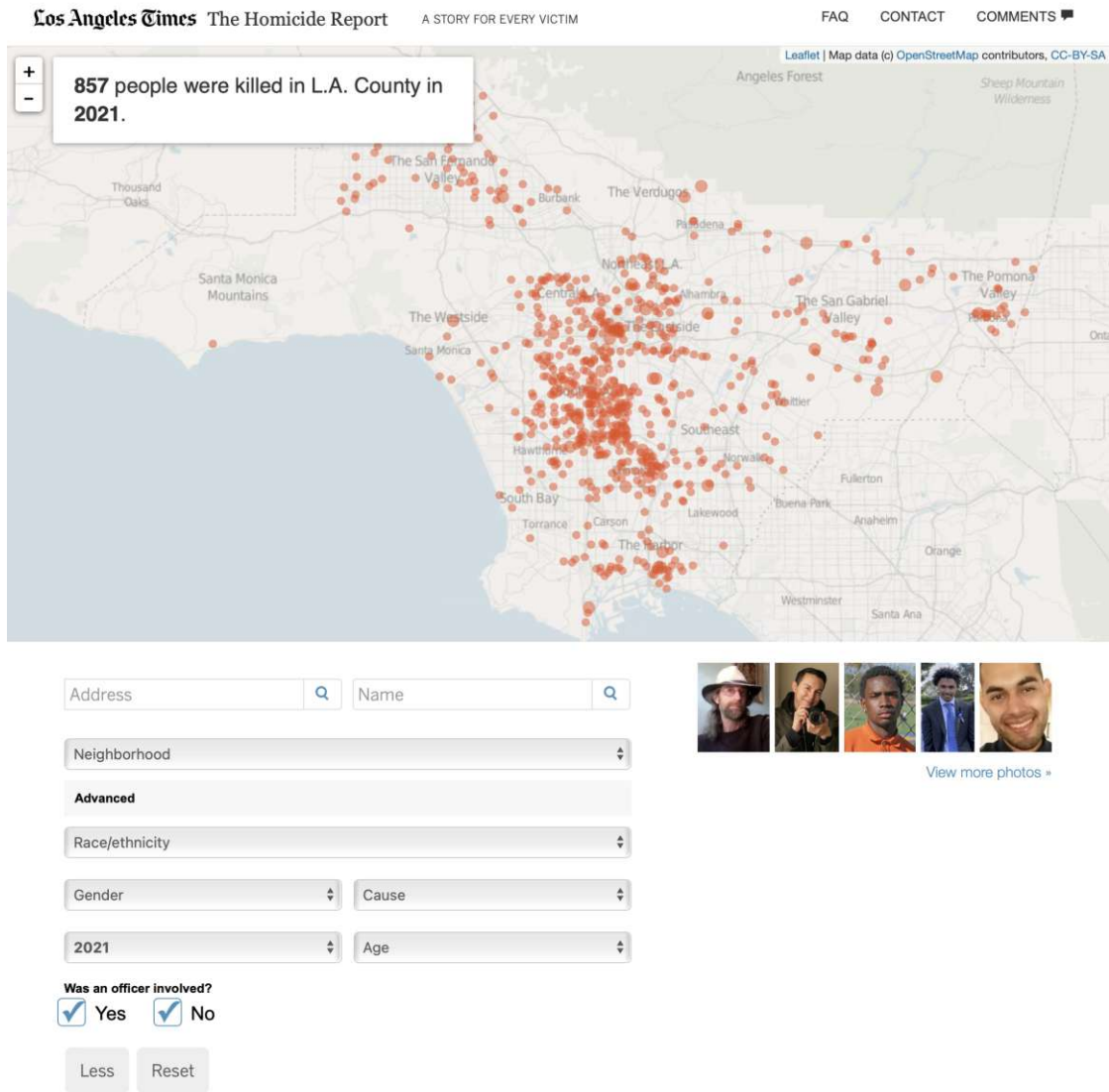
21

Figure 5.1: Map of The Homicide Report - Los Angeles Times[40]

magnitude of earthquakes.[42] A map showing the earthquake's epicenter also gets attached by the system.[43] Besides these basic facts, the significance is set into a historic context by comparing the data to previous earthquakes in the area. Because it only takes eight minutes for Quakebot to publish a finished story, it outpaces any human journalist.[44] But there is still a manual editorial check necessary for the story to be published.[45] However, experience showed that even this manual review failed to prevent the publication of false reports on earthquakes that never existed in that area, caused by incorrect input data.[46] Quakebot has been developed in-house and is based on quite simple code. It greatly improved the speed of publication, while also enlarging the organization's news coverage.[47]

In 2016, *The Washington Post* launched a popular AJ in-house solution named "Heliograph". While it has initially been developed for the 2016 Olympics in Rio, it has since also been used for news on other sport events and for election news.[48],[49],[50] The purpose of the bot was to post very short news on social media. Because of the lack of detail in its reports, the system just complemented the in-dept human written stories on the Post's website.[51] For the Olympics, Heliograph "[...] aimed to streamline the news coverage with live information and permanent updates (events schedules, results, medal tallies, reminders, etc.) of the many competitions that take place simultaneously."[52] For example, it sent out reminders on when certain competitions started, announced the winner of the day and briefly reported on the results.[53] This NLG system helped the Post to attract millions of visitors by establishing AJ as a new line of business, resulting in increased subscriber and advertising revenue.[54] The Chinese company *Tencent* also developed an automated news generation system for the 2016 Olympics in Brazil that predominantly reported about the accomplishments of Chinese athletes.[55]

Another example in the sports domain is the AJ solution provider *Narrative Science*. It came up with an algorithm called "Quill" that produces stories on baseball and softball amateur leagues. A unique feature of this solution is the smartphone app that accom-

---

[42]See Miroshnichenko, "AI to bypass creativity. Will robots replace journalists? (The answer is "yes")", pp. 5, 8.

[43]See Latar, "Robot journalism", p. 37.

[44]See Miroshnichenko, "AI to bypass creativity. Will robots replace journalists? (The answer is "yes")", pp. 5, 8.

[45]See Graefe, *Guide to Automated Journalism*, p. 24.

[46]See ibid., p. 25.

[47]See ibid., p. 20.

[48]See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", p. 43.

[49]See Kunert, "Journalists, meet your new colleague algorithm: The impact of automation on content distribution and content creation in the newsroom", p. 140.

[50]See Torrijos, "Automated sports coverage. Case study of bot released by The Washington post during the río 2016 and Pyeongchang 2018 olympics", pp. 1732, 1737.

[51]See ibid., p. 1744.

[52]Ibid., p. 1743.

[53]See ibid., pp. 1738, 1739, 1741.

[54]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 20.

[55]See Latar, "Robot journalism", p. 37.

panies it. This app allows spectators or club members to manually put in the scores of each game. Based on this data, the system automatically generates game reports.[56] A news agency that also started publishing automated news on minor league baseball is the *Associated Press* (AP). Their solution is provided by *Automated Insights*, a major AJ solution provider, with data coming from the Major League Baseball Advanced Media (MLBAM). It allows them to cover all leagues and teams, something that would have never been possible before with the limited capacity of human journalists.[57] When it comes to team sports, the Swedish media company *MittMedia* stands out with its wide coverage of most team sports that are taking place in Sweden. They are operating their AJ system named "Rosalinda" in cooperation with their solution provider *United Robots*. An external provider is responsible for collecting the input data across the country. Besides sports, MittMedia is also running automated reporting on bankrupted companies and on property deals.[58]

Quarterly company earnings reports is another branch that the Associated Press started to automate in 2014. Again, Automated Insights provided the software, named "Wordsmith", for the project.[59] Various aspects are covered by this reports, "Did a company report a profit or a loss? Did it meet, beat, or miss analyst expectations? Did it do better or worse than it did in the previous quarter or a year earlier?"[60] Overall, the project was a great success for the AP. It allowed them to increase their company earnings report output by more than ten times[61],[62], while containing fewer errors compared to human written stories[63]. More importantly, this automation freed up 20 percent of their journalist's time that could be used for more complex and creative tasks in the newsroom, without resulting in job loss.[64] Because the earnings season is a quarterly event that occurs four times a year, it was a very time-intensive cyclic task for the reporters before the automation was in place.[65] Additionally to providing their readers with more earnings reports in their home country, the U.S., the AP could also expand their coverage internationally to Canada and the EU.[66] *Forbes* started even earlier, since 2012, to automatically produce company earnings previews for their readers with great success. This expanded their audience and increased advertising revenue due to increased site traffic.[67]

---

[56]See Zydorek, "Automatisierte Contentproduktion bei Nachrichten – Automated Journalism", pp. 98–100.

[57]See Galily, "Artificial intelligence and sports journalism: Is it a sweeping change?", p. 49.

[58]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 15.

[59]See Graefe, *Guide to Automated Journalism*, pp. 20, 21.

[60]Lecompte, *Automation in the Newsroom. How algorithms are helping reporters expand coverage, engage audiences, and respond to breaking news. URL: https://niemanreports.org/articles/automation-in-the-newsroom/*.

[61]See Miroshnichenko, "AI to bypass creativity. Will robots replace journalists? (The answer is "yes")", p. 9.

[62]See Graefe, *Guide to Automated Journalism*, p. 29.

[63]See ibid., p. 29.

[64]See ibid., p. 29.

[65]See Linden, "Algorithms for journalism: The future of news work", p. 66.

[66]See Graefe, *Guide to Automated Journalism*, p. 29.

[67]See ibid., p. 30.

There has been another AJ solution called "Valtteri", launched by the *Immersive Automation* team in Finland in 2017. This algorithm's purpose was to produce election news on the Finnish municipal elections in three different languages (Finnish, Swedish, and English). It also allowed for personalization by the readers and helped them to search for articles according to their preferences.[68]

Liu et al. state a very interesting use case for automation in the work routine of news agencies.[69] International news agencies like *Reuters* or the *Associated Press* "[...] command an extensive network of journalists that cover various news topics in many countries, and provide subscription-based services to other news media."[70] Therefore, they have to constantly monitor a large number of channels to quickly pick up on newsworthy events. Because speed is key in this business, enhancing parts of their workflow through automation can provide tremendous value. *Reuters*, an international news agency, is running a bot named "Tracer", which is an end-to-end automation solution based on Twitter data. Extracting all information from tweets also means that the algorithm is not restricted to special topics or domains. By analyzing over twelve million tweets per day, newsworthy events are extracted, summarized and classified regarding their newsworthiness, location, scope, veracity, and novelty. Tracer is able to automate the whole process chain, starting form identifying the news and ending with their dissemination. The fully automatically generated news alerts cover 70 percent of the news daily reported by international news agencies.[71] Fake news can be detected with 60-70 percent accuracy.[72] The system especially shines at detecting unexpected events, like accidents or natural disasters, while outpacing any human journalist. Having said that, journalists can still be faster when it comes to reporting on expected events.[73] This automation solution allows Reuters to report between 8 to 60 minutes earlier than competing global news outlets.[74]

When looking just at the DACH region, there are particular use cases for automated text generation that are commonly used by various newsrooms. The most common ones are election news[75,76,77,78], amateur football[79,80,81], Corona virus newsletters[82,83,84],

---

[68]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 22.
[69]See X. Liu et al., "Reuters Tracer: Toward Automated News Production Using Large Scale Social Media Data".
[70]Ibid., p. 1483.
[71]See ibid., p. 1483.
[72]See ibid., p. 1484.
[73]See ibid., p. 1492.
[74]See ibid., p. 1484.
[75]See interview transcript B, question 1.
[76]See interview transcript D, question 1.
[77]See interview transcript F, question 1.
[78]See interview transcript G, question 1.
[79]See interview transcript A, question 1.
[80]See interview transcript E, question 1.
[81]See interview transcript G, question 1.
[82]See interview transcript A, question 1.
[83]See interview transcript B, question 1.
[84]See interview transcript F, question 1.

and individualized news at municipal level[85,86,87]. There is also a newsroom providing automated stock exchange reports on the companies that performed best and worst within a day.[88] A social media bot automatically reporting on certain topics and automated news on insolvencies are more examples of services currently in operation.[89] One DACH company also reported on automation use cases apart from automated text generation. They are improving the searchability of their internal image and sound archives by using speech-to-text solutions. Last but not least, they are using automated social media monitoring for trend detection and technologies supporting them in data research tasks like detecting anti-constitutional symbols on social media.[90]

5.2   Major Solution Providers

Over the years, several reputable companies emerged that specialised on automated text generation solutions. The biggest ones in the U.S. are *Automated Insights* and *Narrative Science*[91,92,93], which are basically forming a duopoly in their home country[94]. These companies were the first ones entering the market of AJ back in 2010.[95] Both of them emerged from the sports domain by automatically generating game recaps.[96] *Retresco* and *AX Semantics* are well known NLG solution providers from Germany, serving the German-speaking market (DACH region).[97,98,99] Other Germany based NLG providers are *text-on*, *2txt – natural language generation*, and *textOmatic*.[100] For France there is *Syllabs*[101,102] and *LabSense*[103]. Another well established company in the field of NLG solutions is *Arria NLG*, which originates from the United Kingdom.[104,105] A popular provider in China is *Tencent*.[106]

---

[85]See interview transcript B, question 1.

[86]See interview transcript E, question 1.

[87]See interview transcript G, question 1.

[88]See interview transcript A, question 1.

[89]See interview transcript F, question 1.

[90]See interview transcript A, question 1.

[91]See Dörr, "Mapping the field of Algorithmic Journalism", p. 712.

[92]See Graefe, *Guide to Automated Journalism*, p. 19.

[93]See Linden, "Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?", pp. 129, 130.

[94]See ibid., pp. 129, 130.

[95]See ibid., p. 129.

[96]See Graefe, *Guide to Automated Journalism*, p. 17.

[97]See Dörr, "Mapping the field of Algorithmic Journalism", p. 712.

[98]See Graefe, *Guide to Automated Journalism*, p. 19.

[99]See Krawarik et al., *AI.AT.Media. AI and the Austrian Media Sector: Mapping the Landscape, Setting a Course*, p. 94.

[100]See Graefe, *Guide to Automated Journalism*, p. 19.

[101]See Dörr, "Mapping the field of Algorithmic Journalism", p. 712.

[102]See Graefe, *Guide to Automated Journalism*, p. 19.

[103]See ibid., p. 19.

[104]See Dörr, "Mapping the field of Algorithmic Journalism", p. 712.

[105]See Graefe, *Guide to Automated Journalism*, p. 19.

[106]See ibid., p. 19.

Usually, local companies are better at handling their local language when it comes to lexis, syntax and semantics.[107] However, a limiting factor for the journalistic use case is that many of these solution providers are not primarily focused on journalistic content but on e-commerce (e.g. product descriptions) and other use cases like patient summaries and more. This means that they sell their technology to multiple industries.[108,109,110] For e-commerce use cases it is usually all about static descriptions of products.[111] However, especially for journalism, knowledge about facts, the context, and good narrative skills are key for quality content.[112,113] Telling a good story requires an interpretation of the data, thus adding context and causality.[114] This is what makes this domain particularly challenging. At present, ready-made solutions for journalistic content fail to meet all of these requirements.[115] Almost all of the NLG systems commercially available today, but also the ones developed by news organizations themselves, are still rule-/template-based.[116,117,118,119] In general, it is important not to overhype the capabilities of automated text generation systems currently used in real-world applications.[120]

---

[107]See Krawarik et al., *AI.AT.Media. AI and the Austrian Media Sector: Mapping the Landscape, Setting a Course*, p. 94.

[108]See ibid., p. 94.

[109]See Graefe, *Guide to Automated Journalism*, p. 19.

[110]See Elikci, "Integration of Artificial Intelligence Into Austrian Journalism: A Potential Threat to Austrian Journalism?", p. 185.

[111]See interview transcript F, question 1.3.

[112]See Krawarik et al., *AI.AT.Media. AI and the Austrian Media Sector: Mapping the Landscape, Setting a Course*, p. 97.

[113]See interview transcript F, question 1.3.

[114]See interview transcript F, question 1.3.

[115]See Krawarik et al., *AI.AT.Media. AI and the Austrian Media Sector: Mapping the Landscape, Setting a Course*, p. 97.

[116]See Leppänen, Munezero, et al., "Data-Driven News Generation for Automated Journalism", p. 189.

[117]See Leppänen, Tuulonen, and Sirén-Heikel, "Automated journalism as a source of and a diagnostic device for bias in reporting", p. 42.

[118]See Krawarik et al., *AI.AT.Media. AI and the Austrian Media Sector: Mapping the Landscape, Setting a Course*, p. 94.

[119]See interview transcripts.

[120]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 44.

CHAPTER 6

Readers' Perception

A central aspect of AJ is how readers perceive automated news compared to news written by humans. This can be determined by exploring three criteria: credibility, quality, and readability.[1] Credibility refers to the degree to which the news reflects reality and can be described as the perceived trustworthiness, believability, and expertise.[2] Graefe and Bohlken published a meta study encompassing 11 journal articles on readers' perception of automated news.[3] They found that there is no difference in perceived credibility between automated and human-written news. For perceived quality, there was only a small difference, with readers slightly preferring human-made content.[4] Having said that, readers clearly preferred human-written over automated news in terms of readability.[5][6] Consequently, except for readability, differences in the perception of automated and human-written news are insignificant. Another interesting finding is that readers gave higher scores for human-written articles for all three rating parameters when seeing that the attributed author is human, regardless of the actual author. This means that just thinking that the article is human-written was enough to bias the readers towards higher ratings for the article at hand. Given this knowledge, news organizations could be beguiled to refrain from disclosing the authorship of their automatically generated articles. However, at least for credibility and quality, the effect of the article's attributed source was small compared to the effect of its actual message, which is why newsrooms do not have to worry too much about this when their AJ solution generates quality

---

[1] See Graefe and Bohlken, "Automated journalism: A meta-analysis of readers' perceptions of human-written in comparison to automated news", p. 55.
[2] See Y. Wu, "Is Automated Journalistic Writing Less Biased? An Experimental Test of Auto-Written and Human-Written News Stories", p. 5.
[3] See Graefe and Bohlken, "Automated journalism: A meta-analysis of readers' perceptions of human-written in comparison to automated news", p. 51.
[4] See ibid., pp. 55, 56.
[5] See ibid., pp. 55, 56.
[6] See Clerwall, "Enter the Robot Journalist: Users' perceptions of automated content", pp. 525, 526.

29

content. While a huge divergence can be seen for readability, one has to consider the domain AJ is applied to. Sophisticated storytelling is not that important for very factual topics like company earnings reports or weather reports. For topics like amateur football or earthquake alerts, where no news were available before, readers may already be happy to have news on the topic at all.[7] Overall, studies show that readers are hardly able to discern between automated and human-written news when not in knowledge of the author.[8],[9],[10],[11],[12] At least for more simple use cases where AJ is currently deployed, see section *5.1 Practical Applications*, automatically generated news can be considered on par with human journalists. It is important not to generalize these results to more complex topics that are not fact-based and recurring, but require creative human input in form of context-based reasoning, vivid storytelling, investigative research, and more. That said, as NLG algorithms will likely continue to improve, generated text can soon be expected to score better in terms of readability and more complex topics might be covered in the future.[13]

---

[7] See Graefe and Bohlken, "Automated journalism: A meta-analysis of readers' perceptions of human-written in comparison to automated news", pp. 56–58.

[8] See Carlson, "Automated journalism: A posthuman future for digital news?", p. 229.

[9] See Carlson, "The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority", p. 3.

[10] See Kunert, "Journalists, meet your new colleague algorithm: The impact of automation on content distribution and content creation in the newsroom", pp. 141, 142.

[11] See Clerwall, "Enter the Robot Journalist: Users' perceptions of automated content", p. 527.

[12] See Tewari et al., "Perceptions of Human and Machine-Generated Articles", pp. 10, 11.

[13] See Graefe, *Guide to Automated Journalism*, p. 40.

Chances and Opportunities

Automated Journalism paves the way for new opportunities that were not possible before, with human journalists only. It opens up new dimensions and ways of thinking in the news sector.[1] Traditionally, the quantity of stories newsrooms are able to produce is always limited by the amount of human resources at hand. This means that they have to strongly base their decisions on what stories to produce on the newsworthiness of events, in order to attract the maximum number of readers most efficiently.[2] AJ, for the first time, allows news organizations to drastically increase their content output in an economical manner.[3,4,5] While usually not reaching many readers per automated story, the sheer amount of stories adds up for a large audience.[6,7] Automated news generation facilitates the creation of stories that would not have been possible with human manpower only.[8,9] Algorithms just need seconds for what human journalists would need days, which highlights their efficiency.[10] The increased breadth, specificity, and immediacy of automated news can deliver great value to customers for specific domains.[11]

---

[1]See Carlson, "Automated journalism: A posthuman future for digital news?", p. 229.
[2]See Carlson, "The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority", p. 10.
[3]See ibid., p. 11.
[4]See Carlson, "Automated journalism: A posthuman future for digital news?", p. 230.
[5]See Dalgali and Crowston, "Algorithmic Journalism and Its Impacts on Work", p. 3.
[6]See Carlson, "The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority", p. 11.
[7]See Carlson, "Automated journalism: A posthuman future for digital news?", p. 230.
[8]See Dalgali and Crowston, "Algorithmic Journalism and Its Impacts on Work", p. 3.
[9]See Graefe, *Guide to Automated Journalism*, p. 23.
[10]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 11.
[11]See Elikci, "Integration of Artificial Intelligence Into Austrian Journalism: A Potential Threat to Austrian Journalism?", p. 191.

As long as the data allows for it, events can be covered nearly real-time.[12],[13],[14] Being among the first ones reporting on certain events can be very beneficial for the news organization.[15],[16] Moreover, not only does the speed of coverage offer a competitive advantage, but also the fact that automated news generation is predestined for reporting on events other media companies are currently omitting because it is not economically feasible for them.[17],[18] For instance, it would be highly unprofitable to cover every match in amateur football manually, in contrast to automated reporting.[19] Assuming AJ gets deployed on a topic of recurring nature, with low internal variability, it can be considered low-cost in the long term because costs are only high for the system introduction. Once the first article is successfully generated, variable costs for every article following are low due to little manual intervention required.[20],[21] The initial outlay for AJ projects is usually very high, with a lot of work involved in the data and texts. But once this initial phase is overcome, a scaling effect starts to apply. The longer an automated content production system is in use, the more time relief an organization can benefit from. This time savings can help journalists to focus even more on their actual job, which is telling actual stories.[22] However, news automation can only be economically feasible if the data is high-volume and/or high-velocity. Economically feasible means that, over the whole automation system's lifespan, the set-up and maintenance costs must not be higher than the overall income generated by the system's output. High-volume low-velocity scenarios imply a huge story output within a short period of time, for example for election result reporting. Low-volume high velocity scenarios on the other hand result in a constant stream of story output or constantly updated stories like weather reports. In case neither of these two criteria is fulfilled, chances for profitable system operation are low, unless set-up costs are minimal.[23]

There are two main business models for monetizing journalistic content, advertising revenue based and/or subscription based. The first one strives for as much traffic as possible to maximize revenue, while the other one aims to convert as many readers to paying subscribers as possible. Instead of subscriptions, customers could also be

---

[12]See Graefe, *Guide to Automated Journalism*, p. 23.

[13]See Zydorek, "Automatisierte Contentproduktion bei Nachrichten – Automated Journalism", p. 106.

[14]See Utasch and Loebbecke, "Intelligent technologies shaping business models for journalistic content provision: A concept matrix", p. 4.

[15]See Zydorek, "Automatisierte Contentproduktion bei Nachrichten – Automated Journalism", p. 106.

[16]See Utasch and Loebbecke, "Intelligent technologies shaping business models for journalistic content provision: A concept matrix", p. 4.

[17]See Zydorek, "Automatisierte Contentproduktion bei Nachrichten – Automated Journalism", p. 106.

[18]See Kotenidis and Veglis, "Algorithmic Journalism—Current Applications and Future Perspectives", p. 253.

[19]See Kunert, "Automation in sports reporting: Strategies of data providers, software providers, and media outlets", p. 10.

[20]See Utasch and Loebbecke, "Intelligent technologies shaping business models for journalistic content provision: A concept matrix", pp. 4, 6.

[21]See interview transcript G, question 1.3.

[22]See interview transcript F, question 1.3.

[23]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, pp. 35, 36.

charged per piece.[24] AJ makes it possible to produce hyperlocal news, which can be highly beneficial for both business models.[25] More local news allow for selling more niche advertisements.[26],[27] But providing highly relevant local news that address the daily life issues of citizens can also help to convert new readers into paying customers.[28] For instance, news can be tailored to readers by automatically covering topics relevant to their municipality or local football club. This degree of granularity in reporting is an important advantage of news automation, being able to please customers.[29],[30]

Another benefit of algorithms is that they can be less error-prone compared to humans, avoiding "[...] simple mistakes like misspellings, calculation errors, or overlooking facts."[31] It is important to note that this only holds true if the input data is correct and no programming errors exist.[32] Automated Journalism is particularly suited for mobile news consumed via smartphones.[33],[34] Mobile news popularity is growing fast[35] and is characterized through short and concise texts, which is where automated news generation can particularly shine.[36] Furthermore, AJ also allows for the creation of content in multiple languages simultaneously[37],[38],[39],[40],[41],[42], making it possible to expand one's geographic coverage by opening up new markets.[43] Last but not least, AJ enables for the

---

[24]See Utasch and Loebbecke, "Intelligent technologies shaping business models for journalistic content provision: A concept matrix", p. 1.

[25]See interview transcript E, question 1.3.

[26]See Linden, "Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?", p. 131.

[27]See Graefe, *Guide to Automated Journalism*, p. 23.

[28]See Luengo and Herrera-Damas, *News Media Innovation Reconsidered*, p. 130.

[29]See interview transcript E, question 1.3.

[30]See interview transcript B, question 1.3.

[31]Graefe, *Guide to Automated Journalism*, p. 23.

[32]See ibid., p. 23.

[33]See Thurman, Dörr, and Kunert, "When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism's capabilities and consequences", p. 11.

[34]See Weber and Kosterich, "Coding the News: The role of computer code in filtering and distributing news", pp. 313, 314.

[35]See ibid., pp. 313, 314.

[36]See Thurman, Dörr, and Kunert, "When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism's capabilities and consequences", p. 11.

[37]See Elikci, "Integration of Artificial Intelligence Into Austrian Journalism: A Potential Threat to Austrian Journalism?", p. 185.

[38]See Vázquez-herrero et al., *Journalistic Metamorphosis: Media Transformation in the Digital Age*, p. 23.

[39]See Zydorek, "Automatisierte Contentproduktion bei Nachrichten – Automated Journalism", p. 107.

[40]See Graefe, *Guide to Automated Journalism*, p. 19.

[41]See Thurman, Dörr, and Kunert, "When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism's capabilities and consequences", p. 11.

[42]See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", p. 46.

[43]See Zydorek, "Automatisierte Contentproduktion bei Nachrichten – Automated Journalism", p. 107.

personalization of stories to the specific reader or groups of readers.[44,45,46,47,48,49,50,51] Of course, this is only possible with extensive customer profiles created by collecting information about the users' preferences and behavior.[52,53,54,55,56] By combining this user data with the input data for the story, multiple personalized versions of a story can be generated.[57,58] This allows for individualization in terms of tonality, point of view, and specific user interests.[59] For example, there could be stories on a sporting event highlighting different perspectives to account for fans of the winning team as well as fans of the loosing team separately.[60] Moreover, the algorithms can create personalized news feeds to show readers only what they are interested in.[61,62,63] Ultimately, these personalized user experiences can potentially increase customer loyalty[64], which can in turn result in a competitive advantage.[65] What makes news personalization challenging is gathering the data to build up meaningful customer profiles. Compared to well established and accepted big tech companies, this can be especially difficult because of the rising awareness regarding privacy and legislation like the GDPR.[66]

[44]See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", p. 46.

[45]See Carlson, "Automated journalism: A posthuman future for digital news?", p. 231.

[46]See Graefe, *Guide to Automated Journalism*, p. 26.

[47]See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, p. 11.

[48]See Thurman, Dörr, and Kunert, "When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism's capabilities and consequences", p. 10.

[49]See Carlson, "The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority", p. 11.

[50]See Zydorek, "Automatisierte Contentproduktion bei Nachrichten – Automated Journalism", p. 106.

[51]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 47.

[52]See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", p. 46.

[53]See Carlson, "Automated journalism: A posthuman future for digital news?", p. 231.

[54]See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, p. 11.

[55]See Carlson, "The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority", p. 11.

[56]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 47.

[57]See Carlson, "Automated journalism: A posthuman future for digital news?", p. 231.

[58]See Carlson, "The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority", p. 11.

[59]See Zydorek, "Automatisierte Contentproduktion bei Nachrichten – Automated Journalism", p. 106.

[60]See Graefe, *Guide to Automated Journalism*, p. 26.

[61]See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, p. 11.

[62]See Carlson, "Automated journalism: A posthuman future for digital news?", p. 231.

[63]See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", p. 46.

[64]See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, p. 11.

[65]See Zydorek, "Automatisierte Contentproduktion bei Nachrichten – Automated Journalism", p. 106.

[66]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 47.

CHAPTER 8

Risks and Challenges

Automated news generation not only comes with advantages and benefits, but also with several risks and challenges. It is crucial that these challenges get addressed when a news organization deploys its own AJ solution.

8.1   Professional Journalistic Values

The journalistic profession is made up of several fundamental journalistic values, influencing the way news is produced as well as how it is perceived in society.[1] These values can differ across cultures and organizations.[2] However, a common set of values is described in the literature. The most important ones are: *truth, transparency, credibility, trust, public interest, accountability, impartiality,*[3],[4],[5],[6] and *ethics*[7],[8]. *Truth* is the most consistent one and is accepted by journalists around the world.[9] The values *transparency, truth* and *trust* are closely intertwined with *credibility. Transparency* means not to withhold any information and to disclose the sources and methods used. *Credibility* comes along

---

[1]See Komatsu et al., "AI Should Embody Our Values: Investigating Journalistic Values to Inform AI Technology Design", p. 2.

[2]See ibid., p. 2.

[3]See ibid., p. 2.

[4]See Schapals and Porlezza, "Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions", p. 19.

[5]See Linden, "Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?", p. 127.

[6]See Luengo and Herrera-Damas, *News Media Innovation Reconsidered*, pp. 141, 142.

[7]See Schapals and Porlezza, "Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions", p. 19.

[8]See Linden, "Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?", p. 127.

[9]See Komatsu et al., "AI Should Embody Our Values: Investigating Journalistic Values to Inform AI Technology Design", p. 2.

35

with truth, accuracy, trust, as well as the reputation perceived by the public.[10] Accuracy requires non misleading fact-based reporting.[11] *Accountability* means that journalists and public institutions should be held accountable for their actions, hence encouraging them to act responsibly.[12],[13] Another important value that is equally important both for the readers and journalists is *originality*, meaning that a story should differentiate itself from others, for example by covering unique perspectives or by being the first one reporting on a topic.[14]

All values considered, one must ask how automated systems might affect these fundamental values. When it comes to *truth*, AJ can have positive as well as negative implications. It can facilitate truth by checking facts automatically. A negative implication for truth is that the algorithm is at risk of spreading misinformation when the input data is incorrect, which would require some kind of human verification to compensate for. Misinformation could also be induced by different kinds of bias, see section *8.9 Risk of Bias*. Furthermore, a shortcoming of algorithms is that, while being very good at identifying patterns, they are currently not able to interpret them in a meaningful way. A profound understanding of the context of news events would be necessary to accomplish this. A minimal requirement for trust and transparency is to disclose the software as the author.[15] The impact of automated news generation on the *public interest* and *impartiality* is also twofold. On the one hand, AJ could help to identify and/or create newsworthy stories that are of public interest and to alleviate bias of journalists themselves by automatic review processes. On the other hand, the algorithm's output, als mentioned above, could be biased itself and there are shortcomings regarding the incorporation of contextual information.[16] *Originality* in the lens of news automation is also tricky because while newsworthiness detection algorithms could help to identify new story ideas, unreported events, or new perspectives on events, they rely on social media and many people talking about an event does not necessarily mean it will make an original story. This kind of systems also run the risk of restricting journalists' thinking, if they rely too much on the system's output. All in all, AJ solutions can have a positive and/or a negative impact on each professional journalistic value, thus implying a duality.[17] There is a strong overlap between some of these values, which is why they should be best regarded as a holistic *value ecosystem* for best practise journalism.[18] Ultimately, these values can be understood as the ideology of journalism that can be defined as "[...] a collection of 'values, strategies and formal codes' that characterises journalism and is shared by most members of the occupation as

---

[10]See Komatsu et al., "AI Should Embody Our Values: Investigating Journalistic Values to Inform AI Technology Design", p. 5.

[11]See Leppänen, Munezero, et al., "Data-Driven News Generation for Automated Journalism", p. 189.

[12]See Komatsu et al., "AI Should Embody Our Values: Investigating Journalistic Values to Inform AI Technology Design", p. 7.

[13]See Luengo and Herrera-Damas, *News Media Innovation Reconsidered*, pp. 141, 142.

[14]See Komatsu et al., "AI Should Embody Our Values: Investigating Journalistic Values to Inform AI Technology Design", p. 8.

[15]See ibid., p. 6.

[16]See ibid., p. 7.

[17]See ibid., p. 8.

[18]See ibid., pp. 2, 9.

an active practice that is continually negotiated."[19] All of these ideal-typical journalistic values have to be considered when designing AJ solutions to make sure they get embraced by the system rather than undermined, which is very challenging to achieve.[20] The responsibility for achieving this goal is distributed across all parties involved in building and operating news automation systems. This includes journalists as well as the system's designers, programmers, and data providers.[21]

8.2   Potential of Job Loss

When it comes to the potential of job loss with AJ on the rise, there are two opinions prevalent in the literature, an optimistic and a pessimistic frame.[22],[23] A concern raised by some critical observers is that algorithms might replace human journalists up to a point where they see a threat for the future of human journalists in the long run.[24] Kim and Kim found, in a survey among 42 chief officers from 24 different newspapers in South Korea, that the employment strategy most likely to be chosen when introducing automated news generation is the reduction of human journalists.[25] While this result cannot be generalized because only South Korea was covered[26], financial difficulties, increasing commercial pressure, and growing profit expectations can be a powerful motivator to cut costs wherever possible[27],[28]. Optimistic researchers, on the other hand, "[...] suggest that those fears are mostly unfounded, pointing out that artificial intelligence and algorithms are only going to enhance journalistic practice in the long run instead of replacing it."[29] A benefit commonly agreed upon in the literature is that the time saved through increased efficiency when using news automation can be used by human journalists to

---

[19]Linden, "Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?", p. 127.

[20]See Komatsu et al., "AI Should Embody Our Values: Investigating Journalistic Values to Inform AI Technology Design", p. 2.

[21]See ibid., p. 3.

[22]See Graefe, *Guide to Automated Journalism*, pp. 34, 35.

[23]See Kotenidis and Veglis, "Algorithmic Journalism—Current Applications and Future Perspectives", p. 247.

[24]See ibid., p. 247.

[25]See D. Kim and S. Kim, "Newspaper companies' determinants in adopting robot journalism", p. 193.

[26]See ibid., p. 193.

[27]See S. Wu, Tandoc Jr, and Salmon, "A Field Analysis of Journalism in the Automation Age: Understanding Journalistic Transformations and Struggles Through Structure and Agency", p. 436.

[28]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 9.

[29]Kotenidis and Veglis, "Algorithmic Journalism—Current Applications and Future Perspectives", p. 247.

work on higher value tasks that require human manpower.[30,31,32,33,34,35,36,37,38,39,40] Examples for such high-value tasks where machines struggle are deeper investigations[41], context-heavy stories[42], investigative research[43,44,45], or conducting interviews[46]. This means that AJ can act as a supplementary toolkit for journalists instead of replacing them.[47] When considering the current application areas for automated news, see section *5.1 Practical Applications*, we can see that these are mostly stories that would not be economically feasible to be covered by a human reporter anyway, which also indicates that the fear of job loss is, at least in the short term, mostly unfounded so far.[48,49] However, a journalist whose only job is to write stories on topics that could indeed be fully automated at this point in time would have to reorient himself towards topics and skills algorithms are not able to cover.[50] As of 2017, only between 9 to 15 percent of a newsrooms duties were found to be suitable for automation by experts.[51] There are several areas where AJ falls behind humans. Algorithms generally lack flexibility and cannot draw conclusions or explain new phenomena.[52] Finding the right interview

---

[30]See Elikci, "Integration of Artificial Intelligence Into Austrian Journalism: A Potential Threat to Austrian Journalism?", p. 185.

[31]See Linden, "Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?", p. 132.

[32]See Schapals and Porlezza, "Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions", pp. 21, 24.

[33]See Sirén-Heikel et al., "Unboxing news automation: Exploring imagined affordances of automation in news journalism", p. 55.

[34]See S. Wu, Tandoc Jr, and Salmon, "Journalism Reconfigured: Assessing human–machine relations and the autonomous power of automation in news production", p. 12.

[35]See S. Wu, Tandoc Jr, and Salmon, "When Journalism and Automation Intersect: Assessing the Influence of the Technological Field on Contemporary Newsrooms", p. 13.

[36]See Zamith and Haim, "Algorithmic Actants in Practice, Theory, and Method", p. 2.

[37]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 20.

[38]See Graefe, *Guide to Automated Journalism*, p. 34.

[39]See Milosavljević and Vobič, "Human Still in the Loop: Editors Reconsider the Ideals of Professional Journalism Through Automation", p. 14.

[40]See interview transcript C, question 2.1.

[41]See Schapals and Porlezza, "Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions", pp. 21, 24.

[42]See S. Wu, Tandoc Jr, and Salmon, "Journalism Reconfigured: Assessing human–machine relations and the autonomous power of automation in news production", p. 12.

[43]See ibid., p. 12.

[44]See Graefe, *Guide to Automated Journalism*, p. 34.

[45]See interview transcript C, question 2.1.

[46]See interview transcript C, question 2.1.

[47]See Schapals and Porlezza, "Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions", p. 23.

[48]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 10.

[49]See Dalgali and Crowston, "Algorithmic Journalism and Its Impacts on Work", p. 3.

[50]See Graefe, *Guide to Automated Journalism*, pp. 34, 35.

[51]See Broussard et al., "Artificial Intelligence and Journalism", p. 8.

[52]See Kotenidis and Veglis, "Algorithmic Journalism—Current Applications and Future Perspectives", p. 252.

partners through investigative research is also something currently only humans can do.[53] Machines also lack the ability to ask questions, for example to conduct interviews.[54][55] One of the most definite distinguishing characteristic between humans and algorithms is creativity. At the moment, machines are far from being able to replicate the creativity the human brain is capable of.[56][57][58][59] Furthermore, the ability to write complex stories that require empathy, poetry, humor, metaphors and contextual understanding separates human journalists from automated news.[60][61][62] This lack of curiosity and imagination makes stories predictable.[63] Also, while algorithms shine at revealing correlation[64][65][66] they fail at establishing causality, meaning that they are not able to answer the "why" behind a phenomenon,[67] "[...] due to the inability of computer systems to analyse events against contextual life-world knowledge."[68] This means that even when the linguistic model is of very high quality the algorithm still lacks real world context,[69] which is holding AJ back from the adoption on more complex topics.[70]

However, while the journalists' ability of writing sophisticated context-based stories encompassing vivid storytelling clearly sets them apart from automated news generation, journalists commonly consider skills like objectivity, factuality, simplification and speed as most important for their profession.[71][72] These are skills where algorithms have their

---

[53] See interview transcript C, question 2.1.

[54] See interview transcript C, question 2.1.

[55] See Kotenidis and Veglis, "Algorithmic Journalism—Current Applications and Future Perspectives", p. 252.

[56] See Latar, *Robot Journalism: Can Human Journalism Survive?*, pp. 25, 26.

[57] See Kotenidis and Veglis, "Algorithmic Journalism—Current Applications and Future Perspectives", p. 252.

[58] See Schapals and Porlezza, "Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions", pp. 18, 22.

[59] See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", p. 44.

[60] See Latar, *Robot Journalism: Can Human Journalism Survive?*, pp. 25, 26.

[61] See Y. Wu, "Is Automated Journalistic Writing Less Biased? An Experimental Test of Auto-Written and Human-Written News Stories", p. 3.

[62] See interview transcript C, question 2.1.

[63] See Linden, "Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?", pp. 132, 134.

[64] See Schapals and Porlezza, "Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions", pp. 18, 22.

[65] See Galily, "Artificial intelligence and sports journalism: Is it a sweeping change?", p. 51.

[66] See Miroshnichenko, "AI to bypass creativity. Will robots replace journalists? (The answer is "yes")", pp. 14, 15.

[67] See Schapals and Porlezza, "Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions", pp. 18, 22.

[68] Sirén-Heikel et al., "Unboxing news automation: Exploring imagined affordances of automation in news journalism", p. 50.

[69] See interview transcript F, question 1.4.

[70] See Sirén-Heikel et al., "Unboxing news automation: Exploring imagined affordances of automation in news journalism", pp. 61, 62.

[71] See Dalen, "The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists", p. 653.

[72] See Graefe, *Guide to Automated Journalism*, p. 35.

strengths. News automation is particularly suited for spot news. For some domains, like financial news, sophisticated storytelling would even be counterproductive because the readers' aim is to absorb the relevant information as fast and straightforward as possible.[73] At least on a basic level, algorithms may soon be able to simulate the vividness of human storytelling good enough, for example by providing the system with a large number of variances for different phrases.[74] In the end, algorithms not necessarily have to write better than humans, but just have to be good enough from the readers' perspective.[75] "By operating beauty-meters that rely on the measuring of human reactions, algorithms will be able to automatically produce more attractive texts and headlines without understanding the concept of beauty (or originality, or style)."[76] All in all, it can be concluded that the risk of overarching automation of journalism is very low. Considering the aforementioned differentiating characteristics of human journalists against the background of the journalistic profession, see section *8.1 Professional Journalistic Values*, there are many opportunities for humans to preserve substantial influence to stay in the loop.[77],[78],[79] Of course, journalists constantly need to strive for ways to keep differentiating themselves from algorithms in order to legitimately sustain their professional roles.[80] They need to know the strengths and weaknesses of AJ solutions as well as their own unique skills in order to be able to deliver the best results possible.[81] Instead of challenging machines in their strengths, human newsroom personell should focus on their unique strengths.[82] It is important for reporters to have a good multidisciplinary education and to be familiar with technologies that can support them in producing even better output.[83] In-dept investigative stories are where journalists can particularly set themselves apart from current algorithms' capabilities. This includes gathering unique information from informants and continuously telling stories in new ways by interpreting the context and deeper meaning of the data at hand.[84],[85] While an increasing deployment of automated text generation solutions in certain domains

---

[73] See Graefe, *Guide to Automated Journalism*, p. 35.

[74] See Miroshnichenko, "AI to bypass creativity. Will robots replace journalists? (The answer is "yes")", p. 12.

[75] See ibid., pp. 3, 14.

[76] Ibid., p. 16.

[77] See Linden, "Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?", p. 136.

[78] See Saad and Issa, "Integration or Replacement: Journalism in the Era of Artificial Intelligence and Robot Journalism", p. 12.

[79] See Schapals and Porlezza, "Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions", p. 21.

[80] See ibid., p. 21.

[81] See Latar, "The Robot Journalist in the Age of Social Physics: The End of Human Journalism?", p. 67.

[82] See Dalen, "The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists", p. 653.

[83] See Latar, *Robot Journalism: Can Human Journalism Survive?*, p. 26.

[84] See Schapals and Porlezza, "Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions", pp. 18, 22.

[85] See S. Wu, Tandoc Jr, and Salmon, "A Field Analysis of Journalism in the Automation Age: Understanding Journalistic Transformations and Struggles Through Structure and Agency", p. 438.

is likely to cause less demand for the routine skills associated with that, it remains to be seen whether this automatically raises the demand for previously described higher value tasks that require human manpower.[86] Therefore, journalists should also put emphasis on acquiring computational skills needed for modern news production. In this regard, data analysis skills are becoming more and more important.[87],[88] Having said that, acquiring computational skills does not necessarily mean that journalists have to learn how to code. Instead, "computational thinking" is what matters most[89], meaning that "[...] journalists should be able to work with structured events, narratives, and narrative abstractions [...]."[90] State-of-the-art AJ solutions can be operated via intuitive and practical interfaces, yet they require abstract problem solving skills regarding the use of language and structuring of events. In other words, journalists should be able to think about their language in a similar way software architects think about code.[91] This can also be described as "[a] new kind of 'meta-editorial' skill that combines the ability to abstract the use of language with the ability to apply that abstraction to crafting language-based editorial products [...]."[92]

In general, news can be divided into two categories, spot/hard news and interpretive/soft news, with hard news being more factual and concise compared to soft news, which are more demanding in terms of storytelling and cognitive skills.[93] Finance and sport news are the domains where automated news generation can, at least for spot news type articles, take over a human journalist's job completely. But also in these domains, there is demand for interpretative/soft stories that currently only humans can provide. In sports reporting for example, reporters can serve readers with reactions, analysis and context surrounding the event, focusing on the *why* and *how* instead of the *what*.[94] However, hard and soft news do not necessarily have to be considered separately. Editorial work can also be a mix of the two. In this case, algorithms can provide the body of the story containing all the hard facts, with human journalists building on top of that to add context and additional details.[95] Besides the question of whether the introduction of AJ could lead to job loss, what is certain is that it will create demand for new jobs. Apart from the jobs that are required to build and maintain this kind of software solutions,

[86] See Dalen, "The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists", pp. 655, 656.
[87] See Thurman, Dörr, and Kunert, "When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism's capabilities and consequences", p. 3.
[88] See S. Wu, Tandoc Jr, and Salmon, "A Field Analysis of Journalism in the Automation Age: Understanding Journalistic Transformations and Struggles Through Structure and Agency", p. 437.
[89] See Caswell and Dörr, "Automated Journalism 2.0: Event-driven narratives: From simple descriptions to real stories", pp. 4, 15–17.
[90] Ibid., p. 16.
[91] See ibid., pp. 4, 15–17.
[92] Ibid., p. 17.
[93] See B. Liu and Wei, "Machine Authorship In Situ: Effect of news organization and news genre on news credibility", pp. 6, 7.
[94] See Torrijos, "Automated sports coverage. Case study of bot released by The Washington post during the río 2016 and Pyeongchang 2018 olympics", p. 1735.
[95] See Thurman, Dörr, and Kunert, "When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism's capabilities and consequences", p. 10.

there is also demand for less technical tasks, like writing and updating of news templates, data and knowledge management, parameterization, and more. All of these tasks can be accomplished by tomorrow's journalists.[96],[97]

8.3   Challenges in Newsroom Restructuring

Integrating automated news generation into one's newsroom is a complex endeavor where multiple dimensions have to be considered. According to Kim and Kim, criteria that are most important for C-level executives for deciding on the deployment of AJ are readers' receptivity, governance's outlook, and social trend. On the other hand, sunken costs, journalists' attitudes, and competitors' strategies are considered less important by chief executives.[98] This means that market forecasting and whether customers accept automated news are important priorities. Moreover, the top-management expects a positive cost-benefit ratio that results in a return of investment.[99] Another critical factor is the effectiveness of the technology, or more precisely, its capability of improving performance as well as competitiveness compared to both former routines and external competitors.[100] However, a big problem is the individuals' resistance to organizational change. Employees tend to prefer the status quo because structural change implies stress and uncertainty, as existing routines need to be changed when a new technology gets introduced.[101],[102] A common problem for news organizations is that, despite being largely open for news automation, they do not know where to start and tend to fear machine error and job loss.[103] A system that causes many errors, thus not performing as expected, could result in extra work for journalists to make up for it, which could in turn increase the employees' motivation to leave the company.[104] There are three main concerns that can prevent journalists from endorsing AJ. First, there is the fear of a large number of errors that could be caused by the limitations and imperfections of AJ solutions, thus spoiling journalism quality. Second, some journalists fear a decline in their status as human journalists both within their organization and in society because of damage to journalism's value caused by algorithms.[105] Before the advent of automated news generation only humans were able to write news articles, hence resulting in an exclusive position with social prestige.[106] Third, the potential risk of job loss can be a major

[96] See Saad and Issa, "Integration or Replacement: Journalism in the Era of Artificial Intelligence and Robot Journalism", pp. 9, 10.

[97] See Broussard et al., "Artificial Intelligence and Journalism", p. 8.

[98] See D. Kim and S. Kim, "Newspaper companies' determinants in adopting robot journalism", p. 191.

[99] See ibid., p. 192.

[100] See D. Kim and S. Kim, "Newspaper journalists' attitudes towards robot journalism", p. 9.

[101] See ibid., p. 9.

[102] See S. Wu, Tandoc Jr, and Salmon, "When Journalism and Automation Intersect: Assessing the Influence of the Technological Field on Contemporary Newsrooms", p. 12.

[103] See ibid., p. 15.

[104] See D. Kim and S. Kim, "Newspaper journalists' attitudes towards robot journalism", p. 10.

[105] See ibid., pp. 30, 31.

[106] See ibid., p. 10.

reason against the adoption of this new technology.[107] The fear of unemployment is not specific to the news domain. Concerns regarding job stability are a common phenomenon when new technologies are introduced in an organization. It is very important that the organization's top-management takes their employees' concerns seriously. This means that all of these concerns need to be openly addressed to gain the journalists' trust for a seamless deployment of AJ.[108] Also, an important matter for every successful change project is that chief executives as well as managers are truly committed to the product.[109] The top-management has to believe in the project[110] and make their employees see the value of news automation[111]. To accomplish this, the way of how the added value of AJ is conveyed matters most. For example, the management should tell their journalists how this new technology can free them up from tedious tasks to have more time for more exciting stories. Managers should also make sure that the journalists are using the tool effectively by adhering to best practices, while also having an open ear for their feedback.[112]

8.4 Influence of Company Size

Company size is a huge influencing factor when it comes to the introduction of AJ solutions. Smaller news organizations tend to struggle with a lack of resources, which constitutes an entry barrier. Usually, the required technical skills and expertise is not available in-house.[113],[114],[115] Therefore, external solution providers are often needed.[116],[117] But even if a third-party supplier provides the software solution, extensive fine-tuning in cooperation with the service provider might still be needed to achieve the desired final outcome.[118] Financial constraints are another problem small companies are more prone to than bigger ones.[119],[120] They might lack the budget for initial development and/or running costs.[121] On the other hand, big companies are usually facing more inertia to change, which requires good change management and advocates who really drive the

---

[107]See ibid., pp. 9, 10, 31.

[108]See ibid., pp. 31, 32.

[109]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 16.

[110]See ibid., p. 16.

[111]See S. Wu, Tandoc Jr, and Salmon, "A Field Analysis of Journalism in the Automation Age: Understanding Journalistic Transformations and Struggles Through Structure and Agency", p. 440.

[112]See ibid., p. 440.

[113]See Graefe, *Guide to Automated Journalism*, p. 20.

[114]See S. Wu, Tandoc Jr, and Salmon, "When Journalism and Automation Intersect: Assessing the Influence of the Technological Field on Contemporary Newsrooms", p. 14.

[115]See Fanta, "Putting Europe's robots on the map: automated journalism in news agencies", p. 11.

[116]See Graefe, *Guide to Automated Journalism*, p. 20.

[117]See Fanta, "Putting Europe's robots on the map: automated journalism in news agencies", p. 11.

[118]See Sirén-Heikel et al., "Unboxing news automation: Exploring imagined affordances of automation in news journalism", p. 60.

[119]See S. Wu, Tandoc Jr, and Salmon, "When Journalism and Automation Intersect: Assessing the Influence of the Technological Field on Contemporary Newsrooms", p. 12.

[120]See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, p. 9.

[121]See Fanta, "Putting Europe's robots on the map: automated journalism in news agencies", p. 11.

change.[122] Apart from that, access to data is also a crucial determinant for successful news automation. Especially smaller newsrooms with limited resources have a hard time in developing their own data sets and maintaining them.[123] "Media organizations do not necessarily have the skills or the resources to handle large sets of data securely, verify them and ensure that they do not violate privacy."[124] However, relying on external data providers can also be expensive and, in addition, one might not be able to obtain exclusive rights over this data.[125] Having said that, some data can also be obtained from official sources for free[126], for example election results or communal data, which can be a cost efficient way to gain a foothold in AJ. Another promising opportunity for news companies is partnering with academic institutions and researchers. While differences between journalism ethics and norms and academic peer-reviewed research have to be overcome, this kind of partnerships can drive synergies to achieve joint goals.[127]

8.5 Ethical Requirements

Ethics, a sub-category of philosophy[128], "[...] means deciding what actions are justifiable in the interest of individuals and society, given that many decisions involve values, ends, and procedures that may conflict."[129] On a higher level, AJ can be located between the overlapping fields of media and cyber ethics. On a more detailed level, multiple applied fields of ethics make up for AJ, that are, journalism ethics, information ethics, internet ethics, robo/machine ethics, and computer ethics, see figure 8.1.[130] This overlap is grounded in the fact that automated news generation demands various non-journalistic actors apart from human journalists[132], "[...] e.g. algorithms with delegated agency, media organizations, programmer/service providers of NLG, or data collectors."[133] Another example is data utilization, where the ethical correct approach would be to use only accurate, correct, and objective data, while maintaining data security and privacy.[134] Beside the general values of traditional journalism discussed in section *8.1 Professional Journalistic Values*, ethical responsibilities related to AI also have to be considered. In 2019, an EU expert group formulated seven key requirements for trustworthy AI:

---

[122]See S. Wu, Tandoc Jr, and Salmon, "When Journalism and Automation Intersect: Assessing the Influence of the Technological Field on Contemporary Newsrooms", p. 12.

[123]See Sirén-Heikel et al., "Unboxing news automation: Exploring imagined affordances of automation in news journalism", p. 60.

[124]Ibid., p. 60.

[125]See ibid., pp. 60, 62.

[126]See Fanta, "Putting Europe's robots on the map: automated journalism in news agencies", p. 12.

[127]See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, pp. 9, 10.

[128]See Dörr and Hollnbuchner, "Ethical Challenges of Algorithmic Journalism", p. 2.

[129]Luengo and Herrera-Damas, *News Media Innovation Reconsidered*, p. 15.

[130]See Dörr and Hollnbuchner, "Ethical Challenges of Algorithmic Journalism", pp. 3, 4.

[131]Dörr and Hollnbuchner, "Ethical Challenges of Algorithmic Journalism", p. 4.

[132]See ibid., p. 11.

[133]Ibid., p. 11.

[134]See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", p. 45.
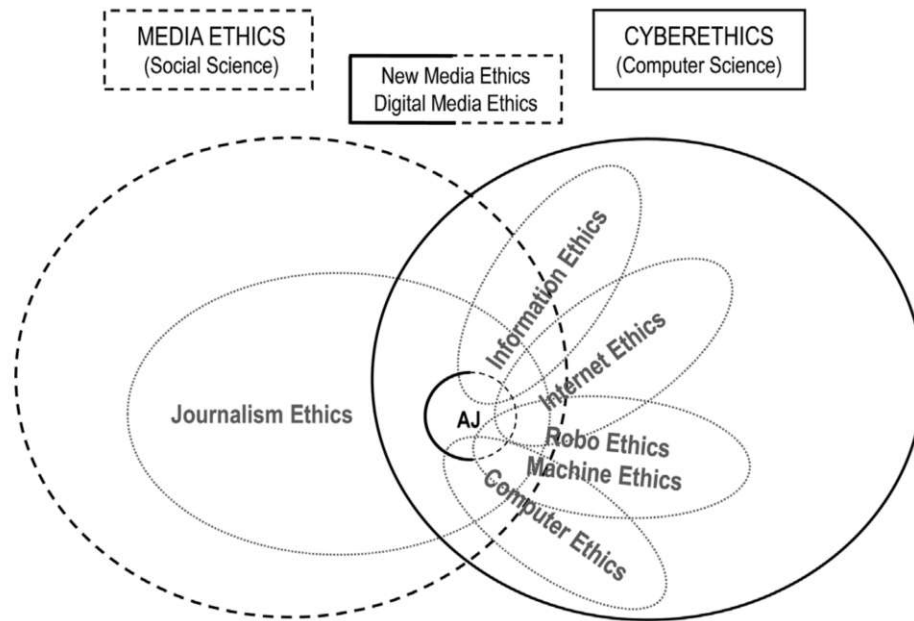
Figure 8.1: Ethical localization of Automated Journalism[131]

"(1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) environmental and societal well-being and (7) accountability."[135] Combining these AI related ethical requirements with the professional journalistic values, see section 8.1, should be a priority that is continuously pursued when designing and operating an AJ solution.

8.6 Algorithmic Transparency and Accountability

Transparency is a very important element for news automation. It is a major influencing factor for reader trust.[136] Ideally, companies provide their readers with the right balance of transparency, which means neither full transparency nor a complete lack of it.[137] However, finding this right balance is challenging. There is a multitude of information readers might want to know. An important category is human involvement, which covers facts like what part of the story is written by an algorithm, editorial goals, details on who created the algorithm, who is held accountable for the produced content, and whether

---

[135] High-Level Expert Group on Artificial Intelligence, *European Ethics Guidelines for Trustworthy Artificial Intelligence (AI)*, p. 24.

[136] See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", p. 44.

[137] See Dalgali and Crowston, "Algorithmic Journalism and Its Impacts on Work", pp. 4, 5.

there is a human review process for automated news.[138] Details regarding the input data is another topic of interest. In case it is not possible to provide readers with the raw data, information like where the data comes from, how it was processed, what parts were used for a particular story and details concerning completeness, accuracy, and timeliness could be of relevance. Some readers might also be interested in the statistical methods and underlying models used in the algorithms.[139],[140] If stories are personalized to the respective reader, the personal data involved is also of importance.[141],[142] The extend to which the reader might be interested in transparency is also dependent on the topic and type of story at hand. Generally speaking, controversial and critical topics, like politics, have a higher tendency for customers to ask for more detailed information regarding the underlying system. The same is true for personalized stories, where readers might be interested in how their story differs from other customers and what personal data was used to generate it.[143]

When it comes to bylines and disclosure for stories created by AJ, different levels of transparency can be found across newsrooms.[144] The most common approach, at least among the well-known news organizations, is *full transparency*[145], which encompasses both a byline crediting the author of the story and a separate full disclosure notice containing the data source used and/or a description of the software vendors. Some newsrooms even state the names of the developers or disclose the full methodology.[146] There is a wide variety for attributing the author of automated news. Some state the software vendor, the programmer, or a human journalist in charge of the system as the author, while others declare the news organization itself as the author. A third popular option is to attribute the name of the algorithm as the author.[147],[148] An alternative to full transparency is *partial transparency*, meaning that there is a byline stating that the article has been automatically generated, but without an additional full disclosure note. A byline only crediting the news organization or a human writer would be an even lower level of transparency. *No transparency* means no byline and further disclosure notes are present at all.[149] However, completely abandoning the byline is not advisable, as "[...] the mere presence of bylines has a measurable impact on readers' perceptions regarding the meaning, authority and credibility of texts [...], while stories without bylines are

---

[138]See Graefe, *Guide to Automated Journalism*, p. 41.

[139]See ibid., p. 41.

[140]See Leppänen, Munezero, et al., "Data-Driven News Generation for Automated Journalism", p. 189.

[141]See Graefe, *Guide to Automated Journalism*, p. 41.

[142]See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", p. 44.

[143]See Graefe, *Guide to Automated Journalism*, pp. 40, 41.

[144]See Montal and Reich, "I, Robot. You, Journalist. Who is the Author?: Authorship, bylines and full disclosure in automated journalism", pp. 8, 9.

[145]See ibid., pp. 12, 13.

[146]See ibid., pp. 8, 12, 13.

[147]See ibid., pp. 8, 12, 13.

[148]See Tandoc Jr, Yao, and S. Wu, "Man vs. Machine? The Impact of Algorithm Authorship on News Credibility", p. 549.

[149]See Montal and Reich, "I, Robot. You, Journalist. Who is the Author?: Authorship, bylines and full disclosure in automated journalism", p. 9.

perceived as dubious and suspicious."[150]

A crucial question for news organizations is what information should be disclosed to the public to achieve the right balance. A good starting point are high level informations like the purpose and goal of the AJ solution and the rationale behind input data selection.[151],[152] Highlighting whether a human editor has cross-checked the article can also be very interesting for readers and potentially increase perceived credibility.[153] Similarly, stating the persons who are in charge of the algorithm can be of interest.[154],[155] What could also be disclosed is whether the software has been created in-house or if the news organization is only responsible for parameterization, with a third-party software vendor involved for the development and maintenance.[156] When it comes to input data, beside basic informations like the data source and privacy implications, communicating the data quality is important and includes information regarding accuracy, completeness, timeliness, limitations, assumptions, and also its representativeness for certain populations.[157] For the model itself its features or variables, along with their associated weights and the rationale behind them, could be disclosed.[158] However, publishing detailed information concerning the uncertainty of the produced output, which could be done by disclosing error margins of the input data[159] or confidence values of the classifiers[160], runs at risk of making the readers think the article is unreliable or even worthless[161]. This is because reaching 100 percent confidence is not possible and readers might interpret scores below 100 as not good enough, no matter if the results are valid from a scientific perspective.[162] Therefore, at least in some instances, sticking to higher level aggregated information for public disclosure is better.[163] But even when the information is released in aggregated form, for things like uncertainty scores it is crucial to provide users with enough contextual information for them to be able to understand the disclosed information[164], which means to "[...] also include a description of the significance of that uncertainty or how that uncertainty can be understood or taken into account by the overall process."[165]

---

[150]Ibid., p. 4.
[151]See Diakopoulos and Koliska, "Algorithmic Transparency in the News Media", p. 8.
[152]See Diakopoulos, "Accountability in Algorithmic Decision-Making: A View from Computational Journalism", p. 12.
[153]See Diakopoulos and Koliska, "Algorithmic Transparency in the News Media", p. 8.
[154]See ibid., p. 8.
[155]See Diakopoulos, "Accountability in Algorithmic Decision-Making: A View from Computational Journalism", p. 12.
[156]See Diakopoulos and Koliska, "Algorithmic Transparency in the News Media", p. 8.
[157]See Diakopoulos, "Accountability in Algorithmic Decision-Making: A View from Computational Journalism", pp. 12–13.
[158]See ibid., pp. 13, 14.
[159]See Diakopoulos and Koliska, "Algorithmic Transparency in the News Media", p. 10.
[160]See Diakopoulos, "Accountability in Algorithmic Decision-Making: A View from Computational Journalism", p. 14.
[161]See Diakopoulos and Koliska, "Algorithmic Transparency in the News Media", p. 13.
[162]See ibid., p. 13.
[163]See Diakopoulos, "Accountability in Algorithmic Decision-Making: A View from Computational Journalism", p. 9.
[164]See Diakopoulos and Koliska, "Algorithmic Transparency in the News Media", pp. 11–13.
[165]Ibid., p. 12.

Source code transparency is also something often discussed in the context of transparency. But making the source code public does not bring benefits to the average user, because technical skills are required to make sense of it.[166],[167]

The overarching goal when disclosing transparency information should be to maintain a good user experience without disturbing the readers who do not care that much, while at the same time providing enough information for the ones who are really interested.[168],[169] This means that media companies have to be aware about the readers' demand for algorithmic transparency.[170],[171] It has been found that few readers actually care about transparency information. Nevertheless, this does not implicate that organizations do not have to care about being transparent.[172] Instead, best ways of disclosing transparency information without causing disturbance have to be identified.[173] One way of doing this would be to provide a link with every automated article for more in-dept information,[174] in addition to a short byline. The rationale behind this approach is to avoid overloading of the common user by revealing in-dept information only on demand.[175] Diakopoulos and Koliska propose a multi-layered *pyramid model* that progressively provides the interested user with more and more information, with only little transparency information on the top-layer and very comprehensive information at the bottom layer.[176] This approach covers "[...] not just a generic general end-user audience but an informed and interested range of actors such as other journalists, critics, activists, or policy makers."[177] It is important to note that disclosing system transparency information does not necessarily have to involve textual content only, but can also include interactivity, especially for stories that involve personalization.[178] This could be realized by allowing users to alter input values to better understand how the algorithm operates. For personalized stories, readers could be allowed to see how the story would look like for other types of readers.[179] Currently, there is no standard regarding transparency disclosure for automated content.[180],[181],[182] It would be beneficial to develop commonly agreed upon best practices for

---

[166] See Diakopoulos, "Algorithmic Accountability: Journalistic investigation of computational power structures", p. 411.

[167] See Diakopoulos, "Accountability in Algorithmic Decision-Making: A View from Computational Journalism", p. 9.

[168] See Diakopoulos, "Algorithmic Accountability: Journalistic investigation of computational power structures", p. 411.

[169] See Diakopoulos, "Accountability in Algorithmic Decision-Making: A View from Computational Journalism", pp. 17, 18.

[170] See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 41.

[171] See Graefe, *Guide to Automated Journalism*, p. 42.

[172] See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 41.

[173] See Graefe, *Guide to Automated Journalism*, p. 42.

[174] See Linden, "Algorithms for journalism: The future of news work", p. 67.

[175] See Diakopoulos and Koliska, "Algorithmic Transparency in the News Media", pp. 12, 13, 15.

[176] See ibid., p. 15.

[177] Ibid., p. 15.

[178] See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 43.

[179] See Diakopoulos and Koliska, "Algorithmic Transparency in the News Media", p. 12.

[180] See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 41.

[181] See Diakopoulos and Koliska, "Algorithmic Transparency in the News Media", p. 2.

[182] See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, pp. 14, 15.

this.[183],[184] A minimum transparency disclosure must indicate whether the article or parts of it were automatically generated.[185] Montal and Reich suggest a byline policy where the software vendor or programmer gets attributed as the author and further information is provided via a separate full disclosure note, including as many of the aforementioned transparency informations as possible.[186] Disclosing whether an algorithm was involved in creating the story can be seen as the first basic layer of system transparency. Linden proposes three more layers, resulting in a four layer transparency model: *disclosure*, *justification*, *explanation*, and *reproduction*. *Justification* means to show that the algorithmic output is acceptable, while an *explanation* elaborates more in-dept on that. In this model, *reproducibility* marks the highest level of transparency by providing the user with enough information to replicate the results himself.[187] In general, the call for transparency may increase when errors occur.[188] Having said that, transparency can also be a competitive advantage for news organizations[189],[190] that helps them to increase credibility[191],[192] and user trust[193],[194]. Understanding how the system in use works can also help readers to better accept the AJ system by building trust. Providing good explainability is crucial in achieving this.[195],[196] Also, not only should transparency be considered towards the readers, but also internally towards the organization itself, in order to be able to explain the origin of errors.[197] This can be facilitated by using software architectures that are build around transparency, for instance by supporting comprehensive logging of the algorithmic state and support for callbacks. Based on such architectures, audit trails could be realized.[198] These kind of software architectures are often referred to as "explainable AI".[199] Currently, many ML-based algorithms in use fail to provide reasons on why a recommendation was made, resulting in a so-called

---

[183]See Diakopoulos and Koliska, "Algorithmic Transparency in the News Media", p. 2.

[184]See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, pp. 14, 15.

[185]See Graefe, *Guide to Automated Journalism*, p. 44.

[186]See Montal and Reich, "I, Robot. You, Journalist. Who is the Author?: Authorship, bylines and full disclosure in automated journalism", pp. 15, 16.

[187]See Linden, "The bright future of news automation", p. 4.

[188]See Graefe, *Guide to Automated Journalism*, p. 45.

[189]See ibid., p. 45.

[190]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 41.

[191]See Graefe, *Guide to Automated Journalism*, p. 45.

[192]See Shin, "Why Does Explainability Matter in News Analytic Systems? Proposing Explainable Analytic Journalism", p. 1060.

[193]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 41.

[194]See Shin, "Why Does Explainability Matter in News Analytic Systems? Proposing Explainable Analytic Journalism", p. 1060.

[195]See ibid., pp. 1050, 1052.

[196]See Shin, "Expanding the Role of Trust in the Experience of Algorithmic Journalism: User Sensemaking of Algorithmic Heuristics in Korean Users", p. 6.

[197]See Linden, "The bright future of news automation", p. 4.

[198]See Diakopoulos, "Accountability in Algorithmic Decision-Making: A View from Computational Journalism", p. 20.

[199]See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, p. 19.

"black box"[200],[201],[202] that hinders the understanding and anticipation of their actions and limits[203]. Two factors that can make a model more interpretable are the number of variables used and the learning technique itself. For example, decision trees are known to deliver interpretable models.[204] It is a fundamental decision that has to be made on what matters more, increased accuracy combined with less or even none interpretability, or more interpretability but less sophisticated and/or accurate models. Only the latter option allows for full transparency.[205]

However, it is important to note that mere technical transparency, no matter how detailed, does not automatically result in *accountability*.[206],[207] This is because accountability also involves the "why", thus demanding explanations on why a particular decision was made, for instance why a system was chosen or who made the decision.[208] Therefore, "[...] accountability requires a more active and involved stance [...]"[209], "[w]hereas transparency is thus passive [..]"[210]. Of course, companies also consider their AJ solutions as a competitive asset, considering them as their intellectual property. This is why complete transparency is often contradicting with trade secrets, which makes it even harder for news organizations to find the optimal balance for transparency disclosure.[211],[212],[213],[214],[215] What must also be considered when it comes to transparency disclosure for automated news generation systems are the associated costs that arise from providing all this information and explanations[216], including tasks such as "[...] data preparation, documentation writing, source code polishing, and benchmark testing."[217]

All in all, reasons for opacity in algorithmic decision making can be summarized by the

---

[200]See Laat, "Big Data and Algorithmic Decision-Making: Can Transparency Restore Accountability?", p. 53.

[201]See ibid., p. 50.

[202]See Hunt and McKelvey, "Algorithmic Regulation in Media and Cultural Policy: A Framework to Evaluate Barriers to Accountability", p. 322.

[203]See Linden, "Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?", p. 128.

[204]See Selbst and Barocas, "The Intuitive Appeal of Explainable Machines", pp. 1110, 1111.

[205]See Laat, "Big Data and Algorithmic Decision-Making: Can Transparency Restore Accountability?", p. 53.

[206]See Ananny and Crawford, "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability", pp. 978, 984.

[207]See Wieringa, "What to Account for When Accounting for Algorithms: A Systematic Literature Review on Algorithmic Accountability", p. 4.

[208]See ibid., p. 4.

[209]Ibid., p. 4.

[210]Ibid., p. 4.

[211]See Leppänen, Munezero, et al., "Data-Driven News Generation for Automated Journalism", p. 189.

[212]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 41.

[213]See Laat, "Big Data and Algorithmic Decision-Making: Can Transparency Restore Accountability?", p. 49.

[214]See Diakopoulos, "Algorithmic Accountability: Journalistic investigation of computational power structures", p. 403.

[215]See Diakopoulos, "Accountability in Algorithmic Decision-Making: A View from Computational Journalism", p. 9.

[216]See Coddington, "Clarifying Journalism's Quantitative Turn", p. 11.

[217]Diakopoulos and Koliska, "Algorithmic Transparency in the News Media", pp. 13, 14.

framework provided by Burrell that consists of three types. First, *intentional opacity*, caused by the aim of protecting one's trade secrets. Second, *illiterate opacity*, which is grounded in the lack of technical skills of most people to read code properly and to understand the design of algorithms. Third, *intrinsic opacity*, meaning that certain ML techniques are opaque by nature.[218,219]

8.7 Risk of Errors

Just like for most technologies, there is always a certain risk for AI systems to produce errors. The cause for an error can be manifold and is usually grounded in a flawed algorithm or in inaccurate input data.[220] When developing this kind of systems, it is important to anticipate rare or unusual events in advance. Also, there has to be outlier detection in place, with the opportunity to escalate an event for human monitoring. A prominent example for errors in automated news in an article from 2015 that accidentally reported a great loss of Netflix's share price despite an actual doubling in price. The reason was a seven-to-one split of the stock that the algorithm failed to account for.[221] This highlights the danger of structural changes in source data, which requires an adequate data management and verification process to prevent this from happening. While these processes can be automated[222], sometimes human validation in form of some degree of editorial oversight is still important to reduce errors[223,224,225,226], because algorithms can only establish correlation but not causality[227,228]. Another advisable strategy for reducing incorrect content is to incorporate the readers by implementing a notification system that allows them to flag certain parts of a story as incorrect or biased for further editorial review.[229]

---

[218] See Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms", pp. 3–5.

[219] See Lepri et al., "Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges", pp. 619, 620.

[220] See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 11.

[221] See Graefe, *Guide to Automated Journalism*, p. 24.

[222] See ibid., p. 43.

[223] See ibid., p. 28.

[224] See S. Wu, Tandoc Jr, and Salmon, "Journalism Reconfigured: Assessing human–machine relations and the autonomous power of automation in news production", p. 10.

[225] See Schapals and Porlezza, "Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions", p. 21.

[226] See Monti, "Automated Journalism and Freedom of Information: Ethical and Juridical Problems Related to AI in the Press Field", p. 13.

[227] See Graefe, *Guide to Automated Journalism*, p. 28.

[228] See S. Wu, Tandoc Jr, and Salmon, "Journalism Reconfigured: Assessing human–machine relations and the autonomous power of automation in news production", p. 10.

[229] See Ombelet, Kuczerawy, and Valcke, "Employing Robot Journalists: Legal Implications, Considerations and Recommendations", p. 735.

8.8 Data Quality and Availability

Automated news generation is only possible if enough quality data is available. This data has to be structured in order to be machine readable.[230],[231],[232],[233] Structured means tabular data organized in columns and rows like a database.[234] Because AJ is not possible without structured data, newsrooms have to make sure that high quality data is available before investing in the system.[235] Poor data quality is likely to cause erroneous or misleading stories[236],[237], which is why news organizations need to make sure that their databases are reliable, high quality, and well maintained[238],[239]. Good data quality is made up by several factors. First, one has to know who collected the data and how it was collected. It has to be clear how its accuracy was verified and how frequently the data gets updated.[240] Moreover, news companies should ask themselves: "Does the data show what we think it shows, is it limited, or does it have gaps?"[241] Another basic precondition is that all information relevant to the story must be represented in the data.[242] In general, databases for AJ should be treated like any other data and sources used by journalists for their everyday work, meaning that a verification process should be applied.[243] This evaluation has to be a continuous process[244], just like data maintenance is a continuous task that accompanies news automation[245]. For example, without continuously managing a data set, company names from earnings reports would become outdated over time, resulting in incorrect stories.[246] It should be a top priority for newsrooms to guarantee for the veracity of the data used in their AJ systems.[247] However, sometimes there is no way for news companies to verify the data themselves. In this case, it is an unsolved question who holds ultimate responsibility.[248] It could be "[...] the data provider, the

---

[230]See Dörr, "Mapping the field of Algorithmic Journalism", pp. 708, 709, 716.

[231]See Graefe, *Guide to Automated Journalism*, p. 27.

[232]See Leppänen, Munezero, et al., "Data-Driven News Generation for Automated Journalism", p. 190.

[233]See Fanta, "Putting Europe's robots on the map: automated journalism in news agencies", p. 13.

[234]See S. Wu, Tandoc Jr, and Salmon, "Journalism Reconfigured: Assessing human–machine relations and the autonomous power of automation in news production", p. 11.

[235]See Leppänen, Munezero, et al., "Data-Driven News Generation for Automated Journalism", p. 190.

[236]See Kotenidis and Veglis, "Algorithmic Journalism—Current Applications and Future Perspectives", p. 252.

[237]See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", p. 45.

[238]See Dörr, "Mapping the field of Algorithmic Journalism", p. 709.

[239]See Linden, "Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?", p. 132.

[240]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 42.

[241]Ibid., p. 42.

[242]See interview transcript B, question 1.4.

[243]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 43.

[244]See ibid., p. 43.

[245]See Fanta, "Putting Europe's robots on the map: automated journalism in news agencies", p. 11.

[246]See ibid., p. 11.

[247]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 37.

[248]See Sirén-Heikel et al., "Unboxing news automation: Exploring imagined affordances of automation in news journalism", p. 60.

system provider or the publisher[...]"[249].

It is important to note that good journalism sometimes requires not yet existing data, which might motivate news organizations to build their own data sets for certain topics.[250] What is more, data suitable for AJ needs to be available at the time the topic is journalistically relevant and, ideally, it should be a topic where fast journalistic processing is needed. These two requirements make it even harder to find adequate data and topics for automated news generation.[251] To increase the overall data availability in the news sector, companies should advocate for more open public and private data. Additionally, they should come up with strategies to obtain unique data sets, either by themselves or through partnering, to be able to provide their readers with unique and captivating stories.[252]

8.9 Risk of Bias

An issue that is important to consider in the context of AJ is *algorithmic bias*, which "[...] occurs when the outputs of an algorithm benefit or disadvantage certain individuals or groups more than others without a justified reason for such unequal impacts."[253],[254],[255] Consequently, algorithmic bias may lead to discriminatory results that are not in accordance with norms of justice and equality.[256] However, in order to be classified as bias, this deviation from equality has to be of systematic and repeated nature and not just a random phenomenon.[257],[258],[259] There are three primary causes where algorithmic bias can originate from: input data, methodological approach, or societal/human factors.[260],[261] Input data or, in case of supervised machine learning, training data is a fundamental determinant for potential bias. The input data may be non-representative for the target population[262],[263],[264],[265], thus resulting in potential discrimination against specific de-

---

[249] Ibid., p. 60.

[250] See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, p. 9.

[251] See interview transcript B, question 1.2.

[252] See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 47.

[253] Kordzadeh and Ghasemaghaei, "Algorithmic bias: review, synthesis, and future research directions", p. 1.

[254] See Cramer et al., "Assessing and Addressing Algorithmic Bias in Practice", p. 60.

[255] See Akter et al., "Algorithmic bias in data-driven innovation in the age of AI", p. 4.

[256] See Kordzadeh and Ghasemaghaei, "Algorithmic bias: review, synthesis, and future research directions", pp. 1, 8.

[257] See ibid., p. 8.

[258] See Cramer et al., "Assessing and Addressing Algorithmic Bias in Practice", p. 60.

[259] See Fazelpour and Danks, "Algorithmic bias: Senses, sources, solutions", p. 2.

[260] See Akter et al., "Algorithmic bias in data-driven innovation in the age of AI", p. 5.

[261] See Tal et al., "'End to End' Towards a Framework for Reducing Biases and Promoting Transparency of Algorithmic Systems", p. 3.

[262] See Akter et al., "Algorithmic bias in data-driven innovation in the age of AI", p. 5.

[263] See Fazelpour and Danks, "Algorithmic bias: Senses, sources, solutions" p. 6.

[264] See Kordzadeh and Ghasemaghaei, "Algorithmic bias: review, synthesis, and future research directions", p. 3.

[265] See Laat, "Big Data and Algorithmic Decision-Making: Can Transparency Restore Accountability?", p. 45.

mographic groups[266,267,268]. Having a training data set that is too small can cause bias too.[269] But even if the data itself has been collected accordingly it can still reflect bias measured from the real world.[270] Protected societal groups like racial minorities are often naturally underrepresented because of a lower population.[271] While removing protected attributes like race or gender from input data can help in some cases to alleviate bias[272], the algorithm can still encode the bias present in the data via proxy variables, for example by using the name and postal code instead of a dedicated race attribute. Therefore, it is important to consider the cultural community of the area the data gets collected.[273] The general problem with data bias is that it can be very subtle and hard to tell.[274] Software developers of the AJ solution might not be able to tell whether the input or training data is biased, especially when the data has been collected by another party.

Methodological bias refers to all bias caused during algorithmic processing.[275] For ML approaches, reasons for method bias can be the correlation fallacy, meaning that correlation gets confused with causation, or generic insights resulting from overgeneralization.[276] What should also be avoided is overfitting of a model by not generalizing assumptions that hold only true for the training data and not for the real world.[277] The phenomenon that insensitive proxy variables could be unintentionally used by the model to represent protected sensitive attributes is another example for bias that is grounded in the method.[278]

Societal bias on the other hand is caused directly by humans themselves[279,280,281] and

---

[266] See Fazelpour and Danks, "Algorithmic bias: Senses, sources, solutions", p. 6.

[267] See Tal et al., "'End to End' Towards a Framework for Reducing Biases and Promoting Transparency of Algorithmic Systems", p. 3.

[268] See Fu, Huang, and Singh, "Artificial Intelligence and Algorithmic Bias: Source, Detection, Mitigation, and Implications", p. 53.

[269] See Akter et al., "Algorithmic bias in data-driven innovation in the age of AI", pp. 5, 6.

[270] See Fazelpour and Danks, "Algorithmic bias: Senses, sources, solutions", p. 6.

[271] See Fu, Huang, and Singh, "Artificial Intelligence and Algorithmic Bias: Source, Detection, Mitigation, and Implications", p. 53.

[272] See Tal et al., "'End to End' Towards a Framework for Reducing Biases and Promoting Transparency of Algorithmic Systems", p. 3.

[273] See Diakopoulos, "Accountability in Algorithmic Decision-Making: A View from Computational Journalism", p. 4.

[274] See Danks and London, "Algorithmic Bias in Autonomous Systems.", p. 2.

[275] See Tal et al., "'End to End' Towards a Framework for Reducing Biases and Promoting Transparency of Algorithmic Systems", p. 3.

[276] See Akter et al., "Algorithmic bias in data-driven innovation in the age of AI", p. 6.

[277] See Leppänen, Tuulonen, and Sirén-Heikel, "Automated journalism as a source of and a diagnostic device for bias in reporting", p. 43.

[278] See Tal et al., "'End to End' Towards a Framework for Reducing Biases and Promoting Transparency of Algorithmic Systems", p. 3.

[279] See Akter et al., "Algorithmic bias in data-driven innovation in the age of AI", p. 7.

[280] See Milosavljević and Vobič, "Human Still in the Loop: Editors Reconsider the Ideals of Professional Journalism Through Automation", p. 12.

[281] See Diakopoulos, "Algorithmic Accountability: Journalistic investigation of computational power structures", p. 402.

can therefore also be called human bias[282]. A prominent example are societal biases in the real world, like "[...] disadvantaged populations of different social status, religion, sexual orientation, subcultures, age groups, gender and other social groups."[283] These historic biases can then cause biased data in return.[284] The development of algorithms also requires many decisions to be made by humans, which is why algorithms can be considered as reflections of its human creators' choices.[285],[286] These decisions are influenced by peoples' values and therefore not necessarily objective, hence posing a potential source of human bias.[287],[288] Therefore, whenever an algorithm or model is used outside of the context where its data or inherent system decisions come from, a so called *transfer context bias* can emerge.[289] To avoid this issue, the deployment context has to be similar to the historic context of the underlying data[290] and system decisions.

Bucher states that human editors are often framed as more credible than AJ solutions, despite the fact that humans are prone to bias as well because of their ideology and values.[291] If news organizations put effort into preventing biases from emerging during the whole lifecycle of their AJ systems, AJ can even "[...] make a more reliable claim to factuality, neutrality and objectivity than human journalists."[292] Also, not every bias is necessarily a bad bias, as intentional statistical bias can also be actively used to minimize a known moral societal bias.[293] Unfortunately, there are no clear industry-standard processes for addressing algorithmic bias available yet.[294] Generally, mitigation strategies against bias can be implemented in a pre-processing step directly on the data, in the algorithm in form of an in-processing technique, or in a post-processing step to recalibrate the output accordingly. However, especially societal bias often cannot be addressed this way.[295] For rule-/template-based text generation systems that are based on explicit rules entered by humans, it can be feasible to make sure that no protected attributes, or ones that are correlated with them, are used.[296]

---

[282]See Tal et al., "'End to End' Towards a Framework for Reducing Biases and Promoting Transparency of Algorithmic Systems", p. 3.

[283]Akter et al., "Algorithmic bias in data-driven innovation in the age of AI", p. 7.

[284]See ibid., p. 7.

[285]See Milosavljević and Vobič, "Human Still in the Loop: Editors Reconsider the Ideals of Professional Journalism Through Automation", p. 12.

[286]See Diakopoulos, "Algorithmic Accountability: Journalistic investigation of computational power structures", p. 402.

[287]See Milosavljević and Vobič, "Human Still in the Loop: Editors Reconsider the Ideals of Professional Journalism Through Automation", p. 12.

[288]See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", p. 44.

[289]See Danks and London, "Algorithmic Bias in Autonomous Systems.", p. 4.

[290]See Fazelpour and Danks, "Algorithmic bias: Senses, sources, solutions", p. 8.

[291]See Bucher, "'Machines don't have instincts': Articulating the computational in journalism", p. 927.

[292]Hammond, "From computer-assisted to data-driven: Journalism and Big Data", p. 7.

[293]See Danks and London, "Algorithmic Bias in Autonomous Systems.", pp. 4, 5.

[294]See Cramer et al., "Assessing and Addressing Algorithmic Bias in Practice", p. 59.

[295]See Kordzadeh and Ghasemaghaei, "Algorithmic bias: review, synthesis, and future research directions", p. 4.

[296]See Leppänen, Tuulonen, and Sirén-Heikel, "Automated journalism as a source of and a diagnostic device for bias in reporting", p. 43.

Last but not least, algorithmic bias is closely related to *fairness*.[297,298,299] Because fairness can be very subjective, Tal et al. refer to it as *perceived fairness*[300], which, in context of algorithmic systems, means that "[...] similar people are being treated equally in the classification while still allowing a preferential treatment of individuals in the group."[301] Or more general, "[t]he quality of treating people equally or in a way that is right or reasonable."[302] If algorithmic bias is present, perceived fairness is reduced. However, perceived fairness has a positive influence on peoples acceptance of algorithms and their output.[303] This once again highlights the importance of keeping algorithmic bias to a minimum.

8.10 Risks arising from Mass-Personalization

Like described in chapter *7 Chances and Opportunities*, provided that customer profiles about the users' preferences and behavior are at hand[304,305,306,307,308], AJ enables for the personalization of stories to the specific reader or groups of readers[309,310,311,312,313,314,315,316]. However, on a societal level, personalized news can potentially cause the fragmentation of

[297] See Tal et al., "'End to End' Towards a Framework for Reducing Biases and Promoting Transparency of Algorithmic Systems", pp. 3, 4.

[298] See Kordzadeh and Ghasemaghaei, "Algorithmic bias: review, synthesis, and future research directions", p. 14.

[299] See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", pp. 44–45.

[300] See Tal et al., "'End to End' Towards a Framework for Reducing Biases and Promoting Transparency of Algorithmic Systems", pp. 3, 4.

[301] Ibid., p. 4.

[302] Ibid., pp. 3, 4.

[303] See Kordzadeh and Ghasemaghaei, "Algorithmic bias: review, synthesis, and future research directions", p. 14.

[304] See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", p. 46.

[305] See Carlson, "Automated journalism: A posthuman future for digital news?", p. 231.

[306] See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, p. 11.

[307] See Carlson, "The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority", p. 11.

[308] See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 47.

[309] See Ali and Hassoun, "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities", p. 46.

[310] See Carlson, "Automated journalism: A posthuman future for digital news?", p. 231.

[311] See Graefe, *Guide to Automated Journalism*, p. 26.

[312] See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, p. 11.

[313] See Thurman, Dörr, and Kunert, "When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism's capabilities and consequences", p. 10.

[314] See Carlson, "The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority", p. 11.

[315] See Zydorek, "Automatisierte Contentproduktion bei Nachrichten – Automated Journalism", p. 106.

[316] See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 47.

public opinion[317] by segregating audience members[318]. To counteract this, personalized stories should not only include facts that are interesting to the specific reader, but also include information collectively seen as essential by the newsroom.[319] Of course, this brings up the challenge of balancing "[...] individual desires and interests with the need for a common pool of knowledge across society."[320]

The risk of information overload for the reader is another potential danger associated with automated news generation. This is because of the vast amount of articles that can be generated through automation, which can make it harder for readers to find the content that matters most to them.[321],[322] Therefore, it is crucial for newsrooms to have appropriate techniques for the provision of automated news to their readers. For instance, it is not possible to just publish hundreds of articles on a topic on the front page, assuming that the reader will find what he is looking for. Media companies also have to find the right balance of automated and classic content for their news feed, as storytelling of automated content is usually little exciting because of its very factual style.[323]

8.11 Potential for Misuse

Unfortunately, AJ also comes with the potential for harmful intentional misuse. For example, this technology could be instrumentalized for automated propaganda campaigns.[324] Intentionally biased input data that has been manipulated also runs danger to create misinformation as output. It can be hard for news organizations to find out whether an external data provider, intentionally or unintentionally, provides them with tampered data. This kind of manipulation could be used to influence peoples' opinions towards a hidden agenda.[325] Another potential issue is that the algorithm could be gamed if external parties get to know enough details on the process, which may cause media companies to hesitate from providing too much information on their system to the public, thus having a negative impact on transparency.[326],[327],[328],[329] When a company decides to refrain from full public disclosure, they should at least provide users the option to

---

[317]See Graefe, *Guide to Automated Journalism*, pp. 12, 50.

[318]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 43.

[319]See ibid., p. 43.

[320]Carlson, "Automated journalism: A posthuman future for digital news?", p. 231.

[321]See Kotenidis and Veglis, "Algorithmic Journalism—Current Applications and Future Perspectives", p. 253.

[322]See Thurman, Dörr, and Kunert, "When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism's capabilities and consequences", p. 14.

[323]See interview transcript E, question 1.4.

[324]See Caswell, "Structured Journalism and the Semantic Units of News", p. 18.

[325]See Dalgali and Crowston, "Algorithmic Journalism and Its Impacts on Work", p. 4.

[326]See Linden, "The bright future of news automation", p. 3.

[327]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 41.

[328]See Diakopoulos, "Accountability in Algorithmic Decision-Making: A View from Computational Journalism", p. 9.

[329]See Lepri et al., "Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges", p. 620.

receive a basic transparency report concerning individual decisions that may personally affect them. In any case, official oversight authorities need to have full insight on the system details to preserve accountability.[330] It is important to note that not everyone can game a system just by having full insight on the system details. The ordinary user can only influence attributes that are based on user behavior. Other attributes can only be altered by parties who have direct control over the data, like the data provider itself.[331] Of course, beside the data there could also be manipulation on the algorithm.[332],[333] Again, in case of an external software vendor, this can be hard to detect for news organizations.[334] External parties can also gain information on how an algorithm works by applying *reverse engineering*, which "[...] is the process of articulating the specifications of a system through a rigorous examination drawing on domain knowledge, observation, and deduction to unearth a model of how that system works."[335] But beyond the aforementioned dangers, large scale automated news generation could potentially be used by everyone to influence peoples' individual opinion formation through the so called *popularity bias*. In general, people tend to be afraid to isolate themselves when not being in accordance with the predominant majority opinion. This is why representatives of the perceived majority opinion tend to be more eager to express their believes and opinions publicly, while those who feel they are in the minority are more likely to remain silent.[336] If used inappropriately, AJ could therefore threaten the principle of diversity of opinion by influencing the majority opinion through the sheer amount of articles this technology is capable to produce.[337]

[330]See Laat, "Big Data and Algorithmic Decision-Making: Can Transparency Restore Accountability?", p. 49.
[331]See Diakopoulos, "Accountability in Algorithmic Decision-Making: A View from Computational Journalism", p. 9.
[332]See Linden, "The bright future of news automation", p. 3.
[333]See Dalgali and Crowston, "Algorithmic Journalism and Its Impacts on Work", p. 4.
[334]See Linden, "The bright future of news automation", p. 3.
[335]Diakopoulos, "Algorithmic Accountability: Journalistic investigation of computational power structures", p. 404.
[336]See Habel, *Roboterjournalismus*, p. 56.
[337]See ibid., p. 59.

CHAPTER 9

Legal Implications

Automated news generation poses some legal implications that need to be considered when deploying and running such a system. Companies have to comply with local laws and, if existent, agreements with third-party contractors[1]. Apart from that, there are a number of legal rules that companies based in an EU country have to adhere to when it comes to AI systems:

> "EU primary law (the Treaties of the European Union and its Charter of Fundamental Rights), EU secondary law (such as the General Data Protection Regulation, the Product Liability Directive, the Regulation on the Free Flow of Non-Personal Data, anti-discrimination Directives, consumer law and Safety and Health at Work Directives), the UN Human Rights treaties and the Council of Europe conventions (such as the European Convention on Human Rights)[...]"[2]

However, as AJ becomes more popular and widespread with increasingly advanced algorithms, there might emerge the need for changes in the legal framework in order to encompass news automation explicitly.[3] Areas that are of special interest for AJ are fundamental rights, authorship, copyright, liability for defamation, and explainable algorithmic decision making. The following sections elaborate on these matters.

---

[1]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 42.
[2]High-Level Expert Group on Artificial Intelligence, *European Ethics Guidelines for Trustworthy Artificial Intelligence (AI)*, p. 6.
[3]See Ombelet, Kuczerawy, and Valcke, "Employing Robot Journalists: Legal Implications, Considerations and Recommendations", p. 735.

9.1 Fundamental Rights

AJ and AI systems in general can affect fundamental rights both in a positive or negative way. Therefore, companies should conduct a legal impact assessment regarding their intended technologies before developing them.[4] An expert group set up by the European Commission proposes which families of fundamental rights are especially relevant for AI systems, thus also being relevant for AJ systems. Their suggestion is based on the EU Treaties, the EU Charter of Fundamental Rights, and international human rights law. While the EU Charter only applies to EU law, international human rights law, especially the European Convention on Human Rights, is universally applicable in participating states.[5] According to the expert group, five families of fundamental rights are especially important for trustworthy AI[6]: "respect for human dignity"[7], "freedom of the individual"[8], "respect for democracy, justice and the rule of law"[9], "equality, non-discrimination and solidarity - including the rights of persons at risk of exclusion"[10], and "citizens' rights"[11]. *Respect for human dignity* means that AI systems should never have a negative influence on the the intrinsic worth that every human being has, thus avoiding its repression, compromising, or diminishment.[12] *Freedom of the individual*, in the context of AI, "[...] requires mitigation of (in)direct illegitimate coercion, threats to mental autonomy and mental health, unjustified surveillance, deception and unfair manipulation."[13] To comply with the principle of *equality, non-discrimination and solidarity*, biased algorithmic output has to be avoided[14], see section *8.9 Risk of Bias*. *Citizens' rights* refers to a multitude of rights[15], "[...] including the right to vote, the right to good administration or access to public documents, and the right to petition the administration."[16] However, citizens' rights are unlikely to be affected by automated news generation, which is why it is not that important for the scope of AJ solutions.

Another interesting question is whether automated news is covered by the freedom of expression. Habel evaluates this based on German law.[17] Fundamental rights are not applicable to a software solution itself, because algorithms are no legal entities but legal objects under German law. Fundamental rights only apply to a natural or legal person.[18] However, the output of news automation algorithms can still be covered by the freedom

---

[4]See High-Level Expert Group on Artificial Intelligence, *European Ethics Guidelines for Trustworthy Artificial Intelligence (AI)*, pp. 15, 16.
[5]See ibid., pp. 9, 10.
[6]See ibid., pp. 10, 11.
[7]Ibid., p. 10.
[8]Ibid., p. 10.
[9]Ibid., p. 11.
[10]Ibid., p. 11.
[11]Ibid., p. 11.
[12]See ibid., p. 10.
[13]Ibid., p. 10.
[14]See ibid., p. 11.
[15]See ibid., p. 11.
[16]Ibid., p. 11.
[17]See Habel, *Roboterjournalismus*, pp. 137–143.
[18]See ibid., p. 21.

of expression, that is, when the system's output is ex post explainable.[19] It does not matter whether the output is ex ante explainable or not, meaning that the programmer does not have to know in advance what the exact wording of the final output will look like. The rationale behind this is that the programming of the system's source code represents the phase of the formation of opinion of the programmer.[20] Rule-based systems predominantly used today rely on a specific set of rules and parameters that are directly related to the produced text.[21] Systems that are not ex post explainable are ML-based "black box" approaches like neural nets.[22] According to Habel, it is very unlikely that this type of content is protected by the freedom of expression, because it would not be possible to attribute the generated content to the opinion of a human being.[23] The basic requirement why ex post explainable automated news can be covered by the freedom of expression under German law is that humans are allowed to make use of effective methods of dissemination.[24] Habel also elaborates on the *freedom of the press*, which does not only cover media organizations, but also third-party software vendors that provide the software solution for AJ. This is because in German law the freedom of the press also applies to natural and legal persons who are involved in preparation, production, or distribution of media content.[25]

9.2 Authorship and Copyright

The question of who is the author of news generated by algorithms is not only an ethical one but particularly a legal one[26], as this goes hand in hand with the question of copyright attribution. Because this technology is relatively new, there is very few case law existent so far.[27],[28] There are several possible actors authorship could be delegated to: the software programmer or software company, the data provider supplying the source data, the algorithm itself,[29],[30],[31] the publisher's editor supervising the system

---

[19]See ibid., pp. 137, 138, 139, 140, 142, 143.
[20]See ibid., pp. 142, 143.
[21]See ibid., p. 137.
[22]See ibid., p. 139.
[23]See ibid., pp. 139, 140.
[24]See ibid., p. 143.
[25]See ibid., pp. 98, 101, 102.
[26]See Kunert, "Journalists, meet your new colleague algorithm: The impact of automation on content distribution and content creation in the newsroom", p. 143.
[27]See Díaz-Noci, "Artificial intelligence systems-aided news and copyright: Assessing legal implications for journalism practices", p. 4.
[28]See Ombelet, Kuczerawy, and Valcke, "Employing Robot Journalists: Legal Implications, Considerations and Recommendations", p. 733.
[29]See Weeks, "Media Law and Copyright Implications of Automated Journalism", pp. 91, 92.
[30]See Ombelet, Kuczerawy, and Valcke, "Employing Robot Journalists: Legal Implications, Considerations and Recommendations", p. 733.
[31]See Montal and Reich, "The death of the author, the rise of the robo-journalist: Authorship, bylines, and full disclosure in automated journalism", p. 56.

and selecting the input data, or the publisher of the generated stories[32],[33]. A last option would be to consider automated news as authorless.[34] Attributing authorship to the algorithm itself is highly controversial because it would require legal personhood for the algorithm.[35] Weeks proposes that the choice of authorship attribution will most likely come down to the programmer/programming firm or the data provider, with the latter more likely to be attributed as the author by courts and therefore obtain copyright. The reason for this is that, at least in the U.S., there is more weight given to the economic in lieu of the personal rights of authors when it comes to copyright law. This is also why copyright of all Polaroid photos does not belong to the inventor of Polaroid but to the photographer. The same rationale could be applied to algorithms for automated content generation, meaning that the creator of the algorithm cannot automatically claim all the algorithm's output.[36] Developing complex software is usually a collaborative process involving several people within an organization, which is why it can be argued that the company that developed the software should be ascribed as the author instead of a single person.[37] However, as for now, it is common for many Civil Law countries that authorship is not applicable to software or corporate entities, but only to human creators. Legal literature highlights "[..] the importance of identifying the degree of algorithmic involvement in the creative process and the journalist-machine relations they form."[38] Depending on the intellectual or creative contribution of a human journalist, authorship could be attributed to the human journalist in case the content can be classified as "partly-computer-generated", meaning that the algorithmic output was only intermediary and not the final product.[39],[40] Interestingly, the news organizations pioneering in AJ also have different views on who the author is. Their byline attributions either include the software vendor, the news organization or the algorithm itself. Nevertheless, the majority of them sees their news organization as the author of their automated content.[41] This highlights the *legal-journalistic gap* between legal classification and journalistic practice when it comes to the unique question of authorship for automated news, with most news organizations claiming authorship just as they do for their regular journalistic content.[42]

---

[32]See Ombelet, Kuczerawy, and Valcke, "Employing Robot Journalists: Legal Implications, Considerations and Recommendations", p. 733.

[33]See Montal and Reich, "The death of the author, the rise of the robo-journalist: Authorship, bylines, and full disclosure in automated journalism", p. 56.

[34]See ibid., p. 56.

[35]See Weeks, "Media Law and Copyright Implications of Automated Journalism", p. 92.

[36]See ibid., p. 92.

[37]See Montal and Reich, "I, Robot. You, Journalist. Who is the Author?: Authorship, bylines and full disclosure in automated journalism", p. 9.

[38]Montal and Reich, "The death of the author, the rise of the robo-journalist: Authorship, bylines, and full disclosure in automated journalism", p. 59.

[39]See Habel, *Roboterjournalismus*, pp. 198, 199.

[40]See Díaz-Noci, "Artificial intelligence systems-aided news and copyright: Assessing legal implications for journalism practices", p. 6.

[41]See Montal and Reich, "I, Robot. You, Journalist. Who is the Author?: Authorship, bylines and full disclosure in automated journalism", pp. 9, 12, 13.

[42]See Montal and Reich, "The death of the author, the rise of the robo-journalist: Authorship, bylines, and full disclosure in automated journalism", p. 58.

Montal and Reich's suggestion for correct authorship attribution in media companies' bylines is to state the software vendor or programmer as the author for automated news without additional human involvement. For partly-computer-generated work as discussed above, they suggest to state the human journalist as the author.[43] The resolution of the 2019 AIPPI World Congress on "Copyright in artificially generated works" makes a suggestion based on 32 reports from all over the world, with each report analyzing respective national law. In this resolution the majority considers the investor of the software solution, which can be a natural or legal person, as the original author of content generated by AI.[44]

A remaining question is whether automated news is or should be applicable for copyright at all. At present, copyright acts do not mention automated content production, leaving AJ not explicitly covered in current legal jurisdiction.[45] The aforementioned AIPPI resolution came up with two requirements on this matter. First, "AI generated works should only be eligible for protection by Copyright if there is human intervention in the creation of the work and provided that the other conditions for protection are met."[46] Second, "Originality (as interpreted by national laws) of the generated work resulting from the human intervention should be a condition for the protection by Copyright."[47] The AIPPI resolution also states that a harmonization regarding copyright protection on this topic is desirable.[48] An interesting example of national law is the contribution of the German group to the AIPPI's study. Under German law, only humans are eligible for copyright protection, with machines and corporations being excluded.[49] Also, the group stated that it is unclear whether the programmers' creative inputs for developing the AI does extend to the content generated by the algorithm as well, because there is no case law on this available yet. They concluded that "[u]nder the existing approach, such works need to fulfil the requirements for copyright protection on their own."[50] "Solely AI created works do not deserve copyright protection due to a lack of human involvement and creativity. If, on the other hand, a steering and controlling human influence is present, copyright protection may basically apply."[51] This required human involvement can also be met by a human involved in selecting the input data for the algorithm. However, the input data selection needs to have a "significant influence" to result in a steering influence that is high enough for copyright protection.[52] The Court of Justice of the European Union, when dealing with that question in C-145/10 Eva-Maria Painer v Standard VerlagsGmbH and Others (2011) and in C-604/10 Football Dataco v Yahoo! (2012), "[...] insist[s] in the

---

[43]See ibid., pp. 59, 60.
[44]See Díaz-Noci, "Artificial intelligence systems-aided news and copyright: Assessing legal implications for journalism practices", p. 8.
[45]See ibid., p. 4.
[46]AIPPI World Congress, *APPI. Resolution on copyright in artificially generated works*, p. 2.
[47]Ibid., p. 3.
[48]See ibid., p. 3.
[49]See Freialdenhoven et al., "Do AI generated works qualify for copyright? Summary of the German report to the AIPPI 2019 Study Question on copyright in artificially generated works", p. 28.
[50]Ibid., p. 29.
[51]Ibid., p. 31.
[52]See ibid., p. 30.

decisive intervention of human beings in the final result for a work to be considered copyrightable [...]"[53] While the ultimate question of when the output of a machine qualifies as an original work of a natural person can only be answered by courts[54], the aforementioned information on what may qualify as enough human involvement and creativity provides at least a guideline, although there is no consensus in the legal literature on the limits or criteria for the attribution to a human author[55]. In general, there is a distinction between "computer-assisted-content" and "computer-generated-content", with the latter not having enough human intellectual input to be copyrightable. An example for "computer-assisted" would be content created with the help of word processing programs, which clearly does not raise concerns for a lack of human involvement.[56] In the case of AJ there could also be room for "partly-computer-generated-content", indicating that humans and computers worked together in some way to produce the final work, thus being eligible for copyright protection.[57] Of course, this rises the issue of clear criteria that distinguish "partly-computer-generated" from "(absolutely-)computer-generated". The Swiss jurist Max Kummer came up with a doctrine on that. According to Kummer, it is not yet clear whether the *human predictability* of output can be used as a criterion or whether the *personality of its author* must be recognizable in a produced text. For the current legal doctrine, it is important that a human provided the system with the basic conditions for the content to be generated and that there is some kind of creative selection of the human involved in the process of content generation.[58] According to a decision of the European Court of Justice, for copyright protection to apply "[a] natural person has to express his creativity in an original manner and produce an intellectual creation by the choice, sequence and combination of words (Infopaq International A/S v. Danske Dagblades Forening, CJEU, 2009, § 45)."[59] However, even if the responsible editor can theoretically show creativity by choosing specific input data or questions,[60] "[...] the choice of vocabulary and syntax will be left to the algorithm."[61] It is left to be decided how much human intellectual involvement or creative input is sufficient for copyright law to be applicable.[62],[63]

What has to be considered is that, at present, almost all AJ solutions are rule-based systems instead of way more sophisticated ML-based systems, see section *4.1 Natural Language Generation*. Rule-based systems produce text based on a specific set of rules, defined by humans, which might have implications for the copyright issue regarding

---

[53]See Díaz-Noci, "Artificial intelligence systems-aided news and copyright: Assessing legal implications for journalism practices", p. 7.

[54]See Habel, *Roboterjournalismus*, pp. 185, 186.

[55]See ibid., p. 189.

[56]See ibid., p. 189.

[57]See ibid., pp. 189, 190.

[58]See ibid., p. 190.

[59]Ombelet, Kuczerawy, and Valcke, "Employing Robot Journalists: Legal Implications, Considerations and Recommendations", p. 733.

[60]See ibid., p. 733.

[61]Ibid., p. 733.

[62]See ibid., p. 733.

[63]See Habel, *Roboterjournalismus*, pp. 188, 201.

human involvement and creativity. For example, the 2019 AIPPI resolution covers content generated by AI technology. With ML being a subfield of AI, these considerations may be rather oriented towards sophisticated ML-based systems, which are still in an early experimental stage in the journalistic application for automated news generation. This rises the question to what extent current AJ systems are even able to take over the creative decision making of humans[64],[65]. Habel proposes that, at least under German law, output of ML algorithms that is not ex post explainable does not qualify for copyright protection, because it is no longer a human who sets the basic conditions for text generation and just specifying the topic and desired text length is not enough in that case, since there is no more pre-formulation of text elements, leaving all the creative process of text composition to the machine.[66]

When trying to clarify the issue of copyright protection for content produced by AJ systems, one also has to consider the output's type of work.[67] For newspapers we are dealing with so called *collective works*, which are "[...] composed by many works gathered and structured under the coordination of another (natural or corporate) person."[68] In this context, news organizations are not considered as the author but as the producer of a collective work, because the work is produced under their coordination and due to their investment.[69] The *EU Directive 2019/790 on copyright and related rights in the Digital Single Market* grants press publishers the reproduction right (Article 2 of Directive (EU) 2001/29/EC) and "the exclusive right to authorise or prohibit the making available to the public" (Article 3(2) of Directive (EU) 2001/29/EC) for "the [commercial] online use of their press publications by information society service providers" (Article 15 of Directive (EU) 2019/790), expiring after a period of two years after publication.[70] While this protection is way shorter compared to copyright protection for regular personal creations, it can be considered as sufficient for automated news and is in accordance with the 2019 AIPPI resolution, which considers a shorter protection period for automated content as appropriate.[71] Consequently, even if content created by AJ would not be considered to be eligible for full copyright protection, it would still be protected by EU Directive 2019/790 in EU member states that have implemented it in national law.[72],[73] Another advantage for media companies that arises from this directive is that they are not obliged to provide the authors with "an appropriate share of the revenues" (Article

---

[64]See ibid., p. 190.
[65]See Montal and Reich, "The death of the author, the rise of the robo-journalist: Authorship, bylines, and full disclosure in automated journalism", p. 59.
[66]See Habel, *Roboterjournalismus*, pp. 203–205.
[67]See Díaz-Noci, "Artificial intelligence systems-aided news and copyright: Assessing legal implications for journalism practices", pp. 4, 5.
[68]Ibid., pp. 4, 5.
[69]Ibid., pp. 4, 5.
[70]See ibid., p. 5.
[71]See ibid., p. 5.
[72]See ibid., p. 5.
[73]See Habel, *Roboterjournalismus*, p. 226.

15(2)) if there is no author mentioned in the automatically generated news.[74]

9.3   Liability for Defamation

Like explained in section *8.7 Risk of Errors*, automated news generation always poses the risk of errors or inaccuracies. This also means that there is the risk of defamatory statements that could arise. Consequently, news organizations face the risk of legal actions from potential victims of defamation caused by automated news.[75] In court, the way AJ is implemented can have an influence on the level of fault for news organizations. Media companies using AJ should consider this when it comes to their best practices.[76] The algorithmic origin of news will likely not affect how the truth or falsity of a statement gets assessed in court[77], because "[..] the court adopts an external, reader-centric viewpoint; the intent or status of the writer will not be in question at this point."[78] The same applies to the question of whether the statement has the potential to disparage or discredit the victim in the public opinion.[79] When looking at Austrian law as an example, there has to be differentiated between libel (§ 111 StGB) and slander (§ 297 StGB). While negligence (§ 6 StGB) is enough for libel, actual malice (§ 5 StGB) is required to count as slander. Actual malice would require the publisher to publish information "'with knowledge that the information was false,' or with 'reckless disregard for whether it was false or not.'"[80] To determine whether malice applies, judges have to consider the publishing party's state of mind. This is where automated news differ from regular journalistic content, because it is not the human writer's intent that matters anymore, but responsibility shifts directly to the publisher company.[81] There are two main reasons on why it is unlikely for slander (§ 297 StGB) to apply when it comes to automatically produced news. First, it can be considered highly unlikely that a respected news organization intentionally deploys a news automation solution that produces defamatory statements, which would require the algorithm to be intentionally programmed in a way that it produces false statements. Second, it would be very hard to prove for the defamed individual that it is actual malice[82]. However, media companies could face charges from libel (§ 111 StGB), for which negligence (§ 6 StGB) is sufficient. The risk of charges in this kind of lawsuits, with negligence on the side of media companies, can be found in many jurisdictions[83].

---

[74]See Díaz-Noci, "Artificial intelligence systems-aided news and copyright: Assessing legal implications for journalism practices", p. 5.

[75]See Weeks, "Media Law and Copyright Implications of Automated Journalism", p. 79.

[76]See ibid., p. 79.

[77]See ibid., pp. 79, 80.

[78]Ibid., p. 80.

[79]See ibid., p. 80.

[80]Ibid., p. 80.

[81]See ibid., p. 81.

[82]See Ombelet, Kuczerawy, and Valcke, "Employing Robot Journalists: Legal Implications, Considerations and Recommendations", p. 734.

[83]See Lewis, Sanders, and Carmody, "Libel by Algorithm? Automated Journalism and the Threat of Legal Liability", p. 69.

Of course, if the cause of error is grounded in the algorithm itself, then the developer of the algorithm can be held liable for the damage caused[84]. The same can be said for the data provider, in case erroneous or inaccurate data caused the incident. Having said that, news organizations cannot completely escape responsibility by shifting culpability to third parties. The developer of an algorithm cannot be held accountable for everything the software produces.[85] "Publishers have editorial control over the information that is posted (in e.g. its newspapers), and will therefore be held liable if negligence is shown in its relaying of the information to the public [...]."[86] This means that a newsrooms absence of prudence and monitoring can result in liability.[87] The reason for this is that it can be argued that the news organization should know that their system is not completely foolproof, which is why publishing automated news without any editorial oversight can be considered negligence. For this reason, news organizations should employ some kind of editorial review before publishing.[88] According to Habel, at least for German law, news organizations are not legally obliged to manually review all automatically generated content before publishing, because this kind of proactive review could have a censorship-like effect when considering the amount of articles an AJ solution is potentially able to produce. Therefore, proactive review obligations would have to be limited to options that can be automated according to the state-of-the-art.[89] What also has to be considered in terms of the liability issue for potentially defamatory statements is the domain where the news automation system is used. When AJ gets deployed in domains where the publisher is aware of the increased risk of potential harm to affected persons, the publisher can be held liable for this decision.[90] For example, automated crime reporting causes greater risk compared to sports news in that regard. All in all, it can be concluded that news organizations are well advised in implementing at least some kind of review system to reduce the risk of charges through legal actions of defamed people. Both proactive and reactive ways of reviewing automated news are described in section *8.7 Risk of Errors*.

9.4 Explainable Algorithmic Decision Making

When dealing with AJ systems, EU member states have to consider the EU's General Data Protection Regulation (GDPR), which poses some interesting implications for owners of automated decision systems. An important question in the context of AJ is whether the GDPR mandates explainable algorithmic decision making or, in other words,

---

[84]See Ombelet, Kuczerawy, and Valcke, "Employing Robot Journalists: Legal Implications, Considerations and Recommendations", p. 734.
[85]See ibid., p. 734.
[86]Ibid., p. 734.
[87]See ibid., p. 734.
[88]See Lewis, Sanders, and Carmody, "Libel by Algorithm? Automated Journalism and the Threat of Legal Liability", p. 69.
[89]See Habel, *Roboterjournalismus*, pp. 280, 281.
[90]See ibid., p. 280.

a "right to explanation".[91],[92] The aspect of automated decision making is discussed in GDPR's Articles 13(2)(f), 14(2)(g), 15(1)(h), and 22.[93] The GDPR "[...] requires companies to explain at least 'meaningful information about the logic involved' in automated decision-making, in addition to its significance and envisioned effects (Arts. 13, 14, 15)."[94] There is a debate going on in the legal literature on whether the "meaningful information about the logic involved" demanded by Articles 13-15 requires a system-wide explanation or just concerns explanations on individual decisions. When it comes to Article 22, the debate is about whether the "[...] right to contest an individual algorithmic decision [...] creates an *ex post* right to explanation of an *individual* decision made by an automated system."[95] Also, the granularity of explanations needed is not clearly stated by the GDPR.[96],[97] In the end, these open questions will only be answered conclusively once case law exists on this matter.[98] However, what is for sure is that there is at least some degree of explanation required by the GDPR[99], making news organizations well advised to commit to a high degree of transparency for their AJ solutions and to refrain from "black box" solutions, see section *8.6 Algorithmic Transparency and Accountability*.

---

[91]See Kaminski and Malgieri, "Multi-Layered Explanations from Algorithmic Impact Assessments in the GDPR", pp. 68, 69.

[92]See Selbst and Barocas, "The Intuitive Appeal of Explainable Machines", pp. 1100, 1106, 1107.

[93]See ibid., p. 1106.

[94]Kaminski and Malgieri, "Multi-Layered Explanations from Algorithmic Impact Assessments in the GDPR", p. 69.

[95]Ibid., p. 69.

[96]See ibid., p. 69.

[97]See Shin, "Why Does Explainability Matter in News Analytic Systems? Proposing Explainable Analytic Journalism", p. 1061.

[98]See Linden, "The bright future of news automation", p. 4.

[99]See Selbst and Barocas, "The Intuitive Appeal of Explainable Machines", p. 1107.

Differentiating Characteristics

News organizations constantly strive to maintain and enlarge their number of readers as well as to achieve their revenue targets. This leads to a constant pressure to provide their customers with the best experience possible, with efficient workflows and constant innovation.[1] AJ can be one cornerstone of innovation to differentiate oneself from competitors, thus resulting in a competitive advantage.[2] In general, it is important for news organizations to have a clear focus on what their brand stands for.[3] These topic focus areas of a brand are also important for identifying potentially relevant topics that are particularly well suited for news automation for the specific company at hand. Another important decision criterion for AJ is the primary revenue source of the news organization. Popular business models are based on revenue streams from advertising and/or subscriptions.[4] For example, increasing the number of page impressions by increased news output through automated news generation may be more relevant for advertisement based revenue models than it is for subscription models.[5,6] Therefore, companies have to take their primary revenue model into consideration in order to identify the best use cases of AJ that maximise the competitive potential for the respective company.[7] The degree of regional coverage is also an interesting determinant because of the ability of AJ to produce hyperlocal news. This is especially beneficial for news organizations that can build upon an already existent regional focus in large parts of the country.[8] For AJ to be economically feasible, there is a "[...] need for a minimum quantity of any one 'kind' of

---

[1] See S. Wu, Tandoc Jr, and Salmon, "When Journalism and Automation Intersect: Assessing the Influence of the Technological Field on Contemporary Newsrooms", p. 15.
[2] See interview transcript G, question 1.3.
[3] See interview transcript C, question 4.
[4] See interview transcript D, question 4.
[5] See interview transcript D, question 4.
[6] See interview transcript D, question 4.1.
[7] See interview transcript D, question 4.1.
[8] See interview transcript E, question 4.

story in order to justify the investment in automating its production."[9] This is another reason why many highly relevant and complex topics are not suited for automation and will still require manual writing for a long time to come.[10] Section *8.2 Potential of Job Loss* explains more reasons on this matter.

To differentiate oneself from the competition, media companies should be motivated to innovate and to try new technologies to enhance their workflow, thereby improving both speed and quality of their journalistic output.[11] Building one's own expertise in AJ can also be an advantage over competitors. It allows to build the system according to one's own preferences, for example by adhering to self-imposed ethical guidelines.[12] To accomplish this, news organizations should start to see themselves as technology companies that are invested in state-of-the-art technologies, also apart from AJ. However, especially small media often lack capital for this.[13] Wu et al. propose that the technological field has the ability to redefine journalism in the long run. Evidence for this is the ever more data-centric approach and the trend towards audience centric easily digestible content that can already be seen in journalism. Also, the emerging in-house technologists jobs in journalism may drive significant change over time.[14] Staying ahead of the automation curve will provide companies with a competitive advantage, resulting in more efficient content production and more institutional power by being a leader in the field.[15],[16]

In order to stay ahead of the automation curve, multiple factors have to be considered by news organizations. First and foremost, a best practice approach should be pursued when implementing an AJ solution. Leppänen et al. propose six best practice requirements: "transparency", "accuracy", "modifiability and transferability of the system", "fluency of output", "data availability", and "topicality of news".[17] High *accuracy* means to keep inaccurate output down to a minimum.[18] *Modifiability and transferability* is important to be able to use an AJ system on multiple domains without too much additional adaption costs.[19] *Fluency* of generated content helps to increase reader's satisfaction by providing natural sounding output. *Data availability* also includes an aspiration for high quality data that is well suited for the required topic. *Topicality* can be achieved by covering recent events that are of interest for the readers, for instance localized news that currently lack coverage.[20] When it comes to *transparency*, the Association for Computing Machinery

---

[9] Caswell and Dörr, "Automated Journalism 2.0: Event-driven narratives: From simple descriptions to real stories", p. 16.

[10] See ibid., p. 16.

[11] See interview transcript F, question 4.

[12] See interview transcript A, question 4.

[13] See interview transcript G, question 4.

[14] See S. Wu, Tandoc Jr, and Salmon, "When Journalism and Automation Intersect: Assessing the Influence of the Technological Field on Contemporary Newsrooms", p. 15.

[15] See ibid., pp. 11, 12.

[16] See Utasch and Loebbecke, "Intelligent technologies shaping business models for journalistic content provision: A concept matrix", p. 5.

[17] See Leppänen, Munezero, et al., "Data-Driven News Generation for Automated Journalism", pp. 189, 190.

[18] See ibid., p. 189.

[19] See ibid., pp. 189, 190.

[20] See ibid., p. 190.

US Public Policy Council (USACM) states 7 principles for algorithmic transparency and accountability one should adhere to: "awareness", "access and redress", "accountability", "explanation", "data provenance", "auditability", and "validation and testing".[21] Moreover, media outlets should start to see themselves as their own data hosts/providers instead of relying only on external data providers.[22] At the moment, data handling is often outsourced to third-party providers with news outlets currently showing next to no effort in handling the data themselves.[23],[24] Unfortunately, this also limits the domains where AJ can be applied to what data is externally available. While the reliance on third-party data providers may be convenient and saves on data maintenance costs for news organizations, it also creates an external dependency, which can limit their growth and threaten their autonomy in the long run. Newsrooms can only provide their readers with unique offerings when their data is unique too. However, there can be no unique content when every news outlet uses the same data. This is why news companies should start taking data handling into their own hands to create another competitive advantage in the long term.[25] If news organizations do not own the data, new unexpected competitors could enter the field soon.[26] Service providers for both data and software may start producing their own stories, which is why newsrooms have to make sure that they are able to provide added value over what their intermediaries are able to offer.[27] Beside added value for their readers, another strategy is to cut out on external service providers, for example, where applicable, by gathering and managing the data themselves[28] and by building an in-house solution for their AJ needs.

10.1 Make or Buy Decision

One of the most important and challenging questions to answer for some news organizations is to decide on whether to *make or buy* the software needed for an AJ solution.[29],[30],[31] Generally speaking, the cost-benefit ratio is very important for news outlets.[32],[33] Participant C said that an AJ solution must not cost more than a human journalist for the time saved through the software.[34] What also deters some newsrooms from adopting AJ

---

[21]See Association for Computing Machinery US Public Policy Council (USACM), "Statement on algorithmic transparency and accountability", pp. 1, 2.

[22]See Fanta, "Putting Europe's robots on the map: automated journalism in news agencies", p. 13.

[23]See ibid., p. 13.

[24]See Kunert, "Automation in sports reporting: Strategies of data providers, software providers, and media outlets", p. 9.

[25]See Fanta, "Putting Europe's robots on the map: automated journalism in news agencies", p. 13.

[26]See ibid., p. 13.

[27]See Kunert, "Automation in sports reporting: Strategies of data providers, software providers, and media outlets", p. 13.

[28]See Fanta, "Putting Europe's robots on the map: automated journalism in news agencies", p. 13.

[29]See interview transcript H, question 1.2.

[30]See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, p. 9.

[31]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 38.

[32]See interview transcript C, question 3.

[33]See interview transcript C, question 4.1.

[34]See interview transcript C, question 3.

is the large initial outlay of resources that is necessary.[35] A major advantage of in-house development is that it allows to build knowledge on the topic and to keep it in-house as a competitive advantage.[36] It also adds a lot of flexibility for the way texts are written, simply because code is a very flexible tool that allows to express one's ideas.[37] This direct system access creates independence and enables easier modification.[38] For example, fulfilling one's own ethical guidelines can be easier this way.[39] In-house development also allows for everything to be testable and customizable by yourself and ensures fast response times in case of problems because there is no dependence on third-party providers.[40,41] Participant G noted that, depending on the vendor, it can sometimes take weeks for problems to get solved by external customer support.[42] However, where there is light, there is also shadow. Developing software in-house comes along with maintenance costs that are not to be neglected[43] and ties up additional human resources[44]. Because of this, an appropriate amount of IT resources is required.[45,46,47] But depending on the company size, this can be a barrier for some news organizations, see *8.4 Influence of Company Size*.[48,49] While in-house development can be especially useful for clearly definable use cases like election news, it quickly becomes more complex the broader the intended area of application.[50] Another challenge is the optimization of language models for journalistic application. Most of the big data sets readily available for NLP training are not specific enough for journalistic language and data sets for the domain of journalism are often too small, outdated, or cover just one area of journalistic reporting.[51] Compared to third-party solutions, in-house development often has long lead times of many months before a new tool can go live.[52,53] Companies need to know what their long-term goal is before making a long-term commitment towards developing their own AJ solution, which also requires constant further development and maintenance.[54] It can also be challenging for news organizations to achieve the same usability with their in-house solutions compared to commercial solutions.[55] The software also has to be resilient and fail-safe, including

---

[35]See interview transcript H, question 3.
[36]See interview transcript A, question 1.1.2.
[37]See interview transcript A, question 1.3.
[38]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 39.
[39]See interview transcript A, question 1.1.2.
[40]See interview transcript B, question 1.1.2.
[41]See interview transcript G, question 1.1.2.
[42]See interview transcript G, question 1.1.2.
[43]See interview transcript E, question 1.1.1.
[44]See interview transcript G, question 1.2.
[45]See interview transcript B, question 1.2.
[46]See interview transcript E, question 1.2.
[47]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 39.
[48]See interview transcript E, question 1.2.
[49]See Linden, *News Automation The rewards, risks and realities of 'machine journalism'*, p. 39.
[50]See interview transcript E, question 1.1.2.
[51]See interview transcript A, question 1.2.
[52]See Hansen et al., *Artificial Intelligence: Practice and Implications for Journalism*, p. 9.
[53]See interview transcript B, question 1.2.
[54]See interview transcript E, question 1.2.
[55]See interview transcript A, question 1.4.

advanced error logging with automatic notifications to avoid downtimes, thus increasing costs of development.[56] Last but not least, extensive testing is a major challenge as well, particularly for template-based systems where, according to participant B, a systematic test approach is crucial to reliable test the possible combinations.[57]

Third-party solutions on the other hand also offer some important benefits. First and foremost, they can provide newsrooms with a ready-made user-friendly user interface, which makes them very convenient to use also for people who are not qualified programmers.[58],[59] These external solutions allow for fast results compared to in-house development[60],[61], thus contributing to projects that may be easier to handle for news companies[62]. Another benefit is that the organization's IT resources are less needed.[63] Nevertheless, writing the text generation rules still requires high effort from the newsroom's personnel.[64] There are also third-party solutions available that still leave very much room to work on the input data for media outlets themselves.[65] The downside with external software solutions is that news organizations miss out on the aforementioned benefits of in-house development in the long run. In the end, each news company has to decide on its own what approach, make or buy, suits their company strategy best. They have to weight the benefits and challenges of the two approaches according to their financial and personnel resources at hand against the background of their long-term goal for automated news generation.

10.2 Potential Future Developments

Overall, when taking a look at the big picture, we can see that AJ is only a small segment in the overall automation trend. Many of the processes in the media industry, like information research or content distribution, are already at least partially automated.[66] Participant E expects a large scale adoption of AJ in the future, especially for topics like sports that are present all over the world.[67] A topic that may become popular for AJ is hyperlocal climate reporting.[68] When it comes to possible future developments in AJ, Miroshnichenko expects that algorithms may soon be able to optimize story parameters according to the readers' reactions based on criteria like the number of likes, reposts, comments, or time spent reading.[69] There is also future potential for more hybrid

---

[56] See interview transcript G, question 1.4.
[57] See interview transcript B, question 1.4.
[58] See interview transcript D, question 1.1.1.
[59] See interview transcript D, question 1.3.
[60] See interview transcript E, question 1.1.1.
[61] See interview transcript E, question 1.2.
[62] See interview transcript D, question 1.1.1.
[63] See interview transcript F, question 1.3.
[64] See interview transcript D, question 1.4.
[65] See interview transcript F, question 1.3.
[66] See interview transcript F, question 4.1.
[67] See interview transcript E, question 4.1.
[68] See interview transcript B, question 4.1.
[69] See Miroshnichenko, "AI to bypass creativity. Will robots replace journalists? (The answer is "yes")", pp. 16, 17.

solutions, where a basic text proposal is provided by the algorithm and subsequently enriched with additional information by journalists.[70],[71] What is more, there is always the possibility of large international players to enter the market.[72] This is why it is important as a news organization to at least monitor the field of AJ to have some knowledge in-house, making it easier to anticipate competition and to react accordingly.[73] Increased competitive pressure in the field of news automation might also have positive effects on the availability of suitable data. With increased demand, more quality data might become available, thus creating more opportunities for AJ to be deployed.[74] Each company must explore for itself possible individual benefits of deploying news automation by considering their business model, expectations, audience, and possible domains for application.[75],[76] So far, there are almost exclusively rule-/template-based solutions in real-world use by newsrooms today.[77],[78],[79],[80] However, according to participant F, there is not much potential for innovation left in rule-based systems, except for better language models. Future disruptive innovations will most likely be grounded in ML-based systems that are currently only prototypes for proof of concept, hence not ready for real-world journalistic use yet.[81] All in all, in order to stay competitive, it is very important for news organizations to be sensitive for developments in AJ, including its possible areas of deployment as well as readers' perception and reaction on automated content.[82]

---

[70]See interview transcript A, question 4.1.

[71]See Graefe, *Guide to Automated Journalism*, p. 35.

[72]See interview transcript E, question 4.1.

[73]See interview transcript E, question 4.1.

[74]See interview transcript G, question 4.1.

[75]See interview transcript E, question 4.1.

[76]See interview transcript F, question 4.1.

[77]See Leppänen, Munezero, et al., "Data-Driven News Generation for Automated Journalism", p. 189.

[78]See Leppänen, Tuulonen, and Sirén-Heikel, "Automated journalism as a source of and a diagnostic device for bias in reporting", p. 42.

[79]See Krawarik et al., *AI.AT.Media. AI and the Austrian Media Sector: Mapping the Landscape, Setting a Course*, p. 94.

[80]See interview transcripts.

[81]See interview transcript F, question 4.1.

[82]See Carlson, "Automated journalism: A posthuman future for digital news?", p. 230.

CHAPTER 11

Conclusion

So far, while being important for the future of journalism, AJ is still small-scale with few news organizations active in the field. Structured data has to be available for topics to be suitable for automation. What is more, these topics should be recurring and of cyclic nature, with limited internal variability, and where speed of publication matters. Also, the data needs to be high-volume and/or high-velocity to be economically feasible. Rule-/template-based AJ solutions that are currently predominant in real-world usage are best for describing what happened at a predictable event. This is why as of yet, the most popular domains for automated news are sports, finance, real estate, election results, opinion polls, weather, and traffic news. Common use cases are election news, amateur football, corona virus newsletters, individualized news at municipal level, and company earnings reports. News outlets are confronted with high costs for initial system introduction, but variable costs are comparably low in the long run when the system is used for the same recurring topic, hence implying a scaling effect over time. There is a lot of potential for future disruptive changes from ML-based AJ systems. However, ML-based approaches are still facing major challenges regarding factual correctness and explainability of the output, thus making them not yet suitable for real-world journalistic use. It is remarkable that readers are hardly able to discern between automated and human-written news when not in knowledge of the author, at least for more simple fact-based recurring topics, where AJ is currently deployed. Studies show that differences in perception are mostly only noticeable for readability.

Automated news generation comes along with many great benefits for news organizations. It allows newsrooms to gain more subscribers and to increase advertising revenue by attracting more readers through increased coverage of certain topics that would not be possible with human resources only. Hyperlocal news allow for selling more niche advertisements, but can also increase the number of subscribers when reporting on topics relevant to the daily life of citizens, for instance by covering their local football club or their municipality. For the first time, news organizations are able to drastically increase

their content output in an economical manner, allowing to report on events that would not be economically feasible for other news outlets. Also, increased breadth, specificity, and immediacy of automated news can deliver great value to customers for specific domains. Not only is it possible to cover events in almost real-time, but also certain niches can be covered by this technology. In principle, provided that there is no programming error and input data is correct, algorithms are less error-prone compared to humans, meaning less misspellings, calculation errors, or overlooking facts. Stories can also be created in multiple languages simultaneously, making it possible to expand one's geographic coverage by opening up new markets. Customer loyalty can be increased by the personalization of stories to the specific reader or groups of readers, in terms of tonality, point of view, and specific user interests, provided that meaningful customer profiles are available. Another important benefit is that AJ can free up some of journalist's time for more complex and creative tasks that really require human manpower.

Deploying and operating an AJ system is also associated with several risks and challenges that have to be taken into account. First of all, every party involved in building and operating an AJ system, including journalists, system designers, programmers, and data providers, are responsible to make sure that the system embraces the ideal-typical journalistic values that characterize the journalistic profession, while also adhering to AI related ethical requirements. The fear of job loss after the introduction of automated news generation is mostly unfounded, at least in the short term, because AJ lacks behind when it comes to creativity, context-based reasoning, vivid storytelling, investigative research, and conducting interviews. Moreover, algorithms generally lack flexibility and cannot draw conclusions or explain new phenomena, which holds AJ back from the adoption on more complex topics. Therefore, it is currently deployed for rather simple domains that would not be economically feasible to be covered by human journalists anyway. Nevertheless, increasing commercial pressure and growing profit expectations can be a powerful motivator to cut costs wherever possible. In general, nowadays it is important for journalists to become proficient in computational thinking. When a company wants to deploy AJ, it is crucial that the top management is very committed to the project for the deployment to be a success. This includes to take their employees' concerns seriously. The main take-away regarding the influence of company size is that small companies tend to struggle with providing the required resources for the introduction of automated news generation, while big companies are usually facing more internal inertia to change. Regarding the transparency of these systems, companies have to find the right balance for increased credibility and user trust, with the goal to maintain a good user experience by progressively providing the interested reader with transparency information through a multi-layered pyramid approach. In order to minimize errors, rare or unusual events have to be anticipated and some kind of mechanism to escalate an event for human monitoring should be implemented. Moreover, for news automation to be successful, structured high quality input data has to be available, well maintained, and verified. There are four main types of biases that should be kept to a minimum: data bias, methodological bias, societal/human bias, and the transfer context bias. The latter can be avoided by keeping the deployment context similar to the historic context of the underlying data and system

decisions. News outlets also have to find appropriate techniques for the provision of automated news to their readers to avoid overwhelming them with the sheer amount of content that these automated systems are able to produce. When it comes to personalized automated news, there should always be information included that is collectively seen as essential by the newsroom, in order to prevent the danger of the fragmentation of public opinion by only providing facts interesting to the specific reader. Furthermore, news organizations have to be aware concerning possibilities for both external and internal malicious manipulation of their AJ solutions.

Automated news also come along with unique legal implications. When it comes to fundamental rights, it is important that, under German law, ex post explainable automated news is covered by the freedom of expression and third-party software vendors that are involved in the provision of AJ solutions are covered by the freedom of the press. Regarding the question on who should be attributed as the legal original author, there is no definite consensus in the legal literature yet, with some parties considering the investor of the software solution as the author and others suggesting the programmer/programming firm or the data provider. For copyright protection to apply, steering and controlling human influence with decisive intervention and creativity of human beings is required. However, there is no consensus yet on the actual limits or clear criteria for content to be attributed to a human for copyright protection, which is why it is left to be decided how much human intellectual involvement or creative input is sufficient for copyright law to be applicable. In any case, automated news is protected in EU countries that implemented Article 15 of Directive 2019/790 in national law, granting press publishers a two-year exploitation right for their press publications. To minimize the risk of facing legal actions from the potential victims of defamation caused by automated news, news organizations are well advised to implement some kind of proactive and/or reactive editorial review system, because publishing automated news without any editorial oversight can be considered negligence. News outlets should also be aware that the GDPR requires some degree of explanation of automated decisions (Articles 13(2)(f), 14(2)(g), 15(1)(h), and 22). Therefore, AJ systems are best kept as transparent as possible, avoiding "black box" solutions.

There are several ways news organizations can differentiate themselves in the age of AJ to gain a competitive advantage. It is important for media organizations to be innovative, hence not being afraid to try new technologies that can set them apart from competitors by enhancing their workflow. This includes building one's own expertise in this area, thus also seeing oneself as a technology company that is eager to stay ahead of the automation curve. In general, news companies should try to build on their already established brand focus with their AJ solutions, when it comes to identifying potentially relevant topics for automated news. They also have to tailor their AJ strategy in accordance to their primary revenue model to maximize their competitive potential. Huge potential for gaining competitive advantage lies in news outlets becoming their own data hosts/providers instead of outsourcing everything to external third-party data providers. This way they can strengthen their autonomy and expand their AJ coverage to new domains not accessible before. Also, by coming up with own unique data sets

they can provide their readers with unique and captivating stories. By taking matters into their own hands and cutting out on third-party vendors, news companies can also reduce the risk of external service providers becoming potential new competitors by entering the field on their own. This also includes committing to in-house solutions for AJ, provided that news organizations know they want to be active in the field of AJ for the long term. Of course, companies have to build up enough IT resources beforehand. For news outlets that have limited resources and are not yet sure about their long term strategy when it comes to AJ, third-party solutions are a great way to get a foothold in the market with fast results and comparatively low initial costs. Overall, automated news generation holds great opportunities for news organizations to set themselves apart from the competition, while providing their readers with a new and unique experience.

List of Figures

List of Tables

List of Acronyms

AI Artificial Intelligence. 1, 15, 44, 45, 49, 59, 60, 63, 65, 76

AJ Automated Journalism. xi, xiii, xiv, 1–3, 5, 6, 11, 12, 15, 16, 18–21, 23–26, 29–33, 35–47, 49–78

AP Associated Press. 24

DACH Germany, Austria and Switzerland. xi, xiii, 5, 6, 10, 25, 26

EU European Union. xi, xiii, 3, 24, 44, 59, 60, 63, 65, 67, 77

GDPR General Data Protection Regulation. xi, 34, 67, 68, 77

I-T-O Input-Throughput-Output. 12, 14, 79

IT Information Technology. xi, xiv, 10, 72, 73, 78

L.A. Los Angeles. 21

ML Machine Learning. xi, xiv, 15, 17, 18, 49, 51, 54, 61, 64, 65, 74, 75

MLBAM Major League Baseball Advanced Media. 24

NLG Natural Language Generation. 12, 13, 16–18, 20, 21, 23, 26, 27, 30, 44

NLP Natural Language Processing. 15, 16, 72

SSI Semi-Structured Interview. 6, 7, 9, 10

U.S. United States. 2, 24, 26, 62

USACM Association for Computing Machinery US Public Policy Council. 70, 71

Appendix - Interview Transcripts

Interview A

1. [EN] **Does [company name] currently use intelligent technologies for the automation of workflows, particularly automated news generation? If so, which ones?**

   [DE] **Nutzt [Unternehmen] aktuell intelligente Technologien für die Automation von Workflows, insbesondere zur automatischen Generierung von News? Wenn ja, welche?**

   Generell ist es schwierig zu sagen, wo intelligente Systeme allgemein zum Einsatz kommen, weil es durchaus so ist, dass wir als [Abteilungsname] nicht die einzige Abteilung innerhalb von [Unternehmen] sind, die sich mit Formen der intelligenten Automatisierung auseinandersetzt. Es gibt Anforderungen in verschiedensten Bereichen des journalistischen Arbeitens und der Medienproduktion, wo solche Systeme behilflich sein können, Arbeitsabläufe effizienter oder zumindest effektiver zu gestalten. Wir setzen unterschiedliche Technologien im Bereich der Mediensysteme ein, zum Beispiel zur besseren Durchsuchbarkeit von Bild- und Tonarchiven. Metadaten-Extraktion ist hier ein großes Thema. Häufig kommen hier Speech-to-Text Technologien zum Einsatz, die uns ermöglichen, unsere eigenen Archive besser durchsuchbar zu machen. Gleichzeitig gibt es verschiedene Werkzeuge im Bereich des Social Media Monitorings oder generell Media-Monitoring, die dann auch teilweise auf künstliche Intelligenz setzen, um Trends besser abbilden zu können. Bei uns in [Abteilungsname] arbeiten wir an der Automatisierung von Inhalten. Wir arbeiten vor allem mit Textautomatisierungstechnologien. Dabei geht es um die automatische Erstellung von Texten oder auch der Extrahierung von Texten aus bestehenden Materialien, zur anschließenden Weiterverarbeitung. Wir nutzen unter anderem auch für die datenjournalistische Recherche KI-Verfahren, um verfassungsfeindliche Symbole in Social Media Kanälen zu finden, oder um Satellitenbilder zu analysieren. Also es ist eine große Bandbreite an Dingen, wo man so etwas einsetzen kann. Neben der Sportberichterstattung haben wir ein Projekt, bei dem wir kleine Börsentexte generieren. Hier geht es um die Gewinner und Verlierer des Tages, also welche Unternehmen standen besonders gut oder schlecht da. Zusätzlich haben wir auch einen automatisierten Corona Newsletter, wo wir Texte generieren, welche die Corona Situation in [Land] nochmal zusammenfassen. Das Ganze wird ergänzt durch die typischen Corona Karten, Grafiken und Tabellen, aber mit signifikantem Textanteil.

   1.1 [EN] *(1.1-1.4: If technology for automation is in use)*
   Is the software that is in use developed by your company itself or is it licensed from a third-party supplier?

   [DE] *(1.1-1.4: Wenn Technologie zur Automatisierung in Verwendung)*
   Ist die verwendete Software eine Eigenentwicklung Ihres Unternehmens oder wird sie von einem Drittanbieter lizensiert?

Teils teils. Wir setzen nicht die klassischen Bezahllösungen ein, wie sie zum Beispiel von Retresco oder AX Semantics kommen. Im einfachsten Fall arbeiten wir mit relativ simplen Textvorlagen/Templates, wo entsprechend simpler Regeln Texte und Zahlen eingefügt werden. Auf einem etwas komplizierteren Niveau arbeiten wir mit SimpleNLG, einer Open Source Java Bibliothek, welche die Surface Realisation von natürlicher Sprache erleichtert, womit dann annotiert werden kann welche Funktion ein Satz hat und wir den einzelnen Satzbestandteilen einzelne Funktionen zuordnen können. Das ermöglicht es, deutlich vielfältigere Texte zu schreiben, weil wir uns nicht mehr um einzelne grammatikalische Regeln kümmern müssen. Im Bereich Textzusammenfassung haben wir bereits mit Deep Learning Verfahren gearbeitet, die ein bisschen mehr sind als das klassische Machine Learning. Wir haben aber für uns entschieden, dass dies für den Journalistischen Einsatz, wo es um die Richtigkeit der Informationen geht, bei uns noch nicht einsetzen würden.

1.1.1 [EN] *(If licensed)*
Was it an option to start developing one's own software? Why (not)?
[DE] *(Wenn lizensiert)*
War es eine Option, die Software selbst zu entwickeln? Warum (nicht)?

1.1.2 [EN] *(If self-developed)*
What were the reasons for an in-house development? What benefits and challenges emerged from this?
[DE] *(Wenn Eigenentwicklung)*
Was waren die Gründe für eine Eigenentwicklung? Welche Vorteile und Herausforderungen haben sich dadurch ergeben?
Die Idee bei uns war, dass wir das Wissen im Haus behalten wollten und gleichzeitig auch dieses Wissen zum Thema Textautomatisierung aufbauen wollten. Wir haben uns zudem gewisse ethische Richtlinien auferlegt und haben das Gefühl, dass wir diese am besten erfüllen können, wenn wir die eingesetzte Technologie umfassend verstehen. Ich glaube es gibt sehr viele Textgenerierungslösungen, die unproblematisch sind und wo ethische oder KI-Blackbox Debatten keine Rolle spielen. Wir sehen im Bereich Textautomatisierung auf jeden Fall einen großen strategischen Vorteil für uns und unser Unternehmen und wollten deswegen dieses Wissen selbst aufbauen. Dadurch können wir unser Wissen auch in Bereichen nutzen, wo wir nicht selbst Texte automatisieren, sondern wo wir kritisch über den Einsatz von Textautomatisierung oder allgemein den Einsatz von Algorithmen berichten.

1.2 [EN] **From your current perspective, what is your company's biggest challenge when it comes to making one's own software solutions?**
[DE] **Was ist aus heutiger Sicht die größte Herausforderung für Ihr Unternehmen, wenn es um die Entwicklung eigener Softwarelösun-**

gen geht?

Von der Seite des Machine Learnings betrachtet ist eines der größten Probleme, qualitative Sprachmodelle zu finden. Es geht hier zum Beispiel um vortrainierte Sprachmodelle oder qualitative Trainingsdatensätze, annotierte Artikel, die einem helfen, bestimmte Probleme im Bereich NLP zu lösen. Hier ist es tatsächlich häufig so, dass viele der großen verfügbaren Datensätze nicht spezifisch genug für den Anwendungsfall der journalistischen Sprache angepasst sind. Viele der in diesem Bereich verfügbaren Datensätze, die journalistische Texte codieren, sind häufig viel zu klein, viel zu alt oder nur aus einem einzelnen Bereich des Journalismus. Bei der Optimierung von Sprachmodellen für den Journalismus ist daher noch viel Arbeit zu tun.

1.3 **[EN] What are the biggest benefits that come along with the software solution deployed in your company?**

[DE] Was sind die größten Vorteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?

Unsere eingesetzten Lösungen haben den großen Vorteil, dass sie sehr flexibel in der Art und Weise sind wie wir Texte schreiben, weil wir häufig mit dem Code direkt schreiben und Code generell ein sehr flexibles Werkzeug ist, bestimmte Ideen zum Ausdruck zu bringen.

1.4 **[EN] What are the biggest challenges or downsides that come along with the software solution deployed in your company?**

[DE] Was sind die größten Herausforderungen oder Nachteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?

Andererseits ist es so, dass es nur ein bis zwei Leute im Team mit der nötigen Expertise gibt, die auch Softwareentwickler sind, die damit auch arbeiten können. An dieser Stelle setzen die kommerziellen Lösungen an, weil diese auch Menschen, die keine Softwareentwickler oder Linguisten sind, ermöglichen, Texte aus einfachen Datensätzen zu automatisieren. Hier sind wir noch nicht so weit, dass wir Interfaces für unsere eigene Textautomatisierungslösung anbieten.

2. **[EN]** *(If no automation tools adopted yet)*
Has the introduction of software automation tools ever been considered in your company so far? If so, which ones?

[DE] *(Wenn bisher noch keine Automatisierungstools in Verwendung)*
Wurde die Einführung von softwarebasierten Automatisierungstools in Ihrem Unternehmen jemals in Betracht gezogen? Wenn ja, welche?

2.1 **[EN] Can you think of workflows in a newsroom that would be well suited to be supported or taken over by intelligent technologies?**

[DE] Fallen Ihnen Workflows in der Redaktion ein, die sich eignen würden, von intelligenten Technologien unterstützt oder ersetzt zu werden?

3. [EN] *(If no automated news generation adopted yet)*
   In your opinion, what are the key reasons that kept [company name] from adopting automated news generation so far?

   [DE] *(Wenn bisher noch keine automatisierte Generierung von News in Verwendung)*
   Was sind Ihrer Meinung nach die Hauptgründe, die [Unternehmen] bisher davon abgehalten haben, die automatisierte Generierung von News einzuführen?

4. [EN] **In your opinion, how does [company name] currently differentiate itself from competitors?**

   [DE] **Wie differenziert sich [Unternehmen] Ihrer Meinung nach aktuell von Konkurrenten?**

   Wir gehören zu den Medien, die schon eher viel im Bereich Textautomatisierung machen. Ich glaube was bei uns besonders ist, ist dass wir sehr auf den Prozess achten und auch die ethischen Fragestellungen sehr im Vordergrund stehen und wir hier viel Zeit investieren, um solche Fragestellungen zu klären und hier auch eigene Richtlinien haben. Hier gehen viele andere Unternehmen gefühlt deutlich unbedarfter an die Sache heran.

   4.1 [EN] **With an extensive adoption of automated journalism in the future, how might this current key differentiating characteristics be subject to change?**
   [DE] **Mit einer zukünftig umfassenden Einführung von Automated Journalism, wie könnten sich diese aktuellen Hauptdifferenzierungsmerkmale künftig verändern?**

   Generell muss man sich immer die Fragen stellen, was man damit erzielen möchte. Wenn es einem wirklich nur darum geht mehr Texte zu produzieren, dann gibt es da ganz viele [thematische] Möglichkeiten. Ich glaube viele dieser Ideen im Bereich der Textautomatisierung sind durchaus klickträchtig. Wir müssen uns hier aber auch ganz klar überlegen, inwiefern das unserem Informationsangebot oder Informationsauftrag entspricht. Wir könnten zum Beispiel jeden Tag einen Artikel zu Bahnverspätungen machen. Hier ist aber die Frage, ob das unser Auftrag ist oder ob wir hier nicht einfach klassisch einmal in Jahr recherchieren und dann auch über die Hintergründe hinter den Bahnverspätungen berichten. Ich glaube ganz allgemein setzen wir Automatisierungstechnologien nur an den Stellen ein, wo es auch für Menschen unzumutbar ist jeden Tag den gleichen Text oder einen vergleichbaren Text zu schreiben. Ich glaube was wir auch in Zukunft noch viel mehr sehen werden sind Hybridlösungen, wo der Text nicht zu hundert Prozent aus der Maschine kommt, sondern nur ein Grundgerüst, wo Journalisten dann nochmal die Möglichkeit haben dieses einzuordnen oder mit Zusatzinformationen anzureichern. Zusatzinformationen die wirklich nur Menschen gut liefern können und sich auch nicht aus den Daten ergeben können.

89

Interview B

1. [EN] **Does [company name] currently use intelligent technologies for the automation of workflows, particularly automated news generation? If so, which ones?**

   [DE] **Nutzt [Unternehmen] aktuell intelligente Technologien für die Automation von Workflows, insbesondere zur automatischen Generierung von News? Wenn ja, welche?**

   Was wir bei [Unternehmen] öfter gemacht haben, ist einen Text zu haben, der sich Template-basiert aufgrund der Daten ändert und teilweise umschreibt. Beispielsweise haben wir Geschichten entworfen, wo sich basierend auf der Gemeinde (Standort) die Daten verändert haben. Zum Beispiel Datenfelder, Veränderung von Sätzen oder die Veränderung der Reihenfolge der Sektionen. Das lief alles Template-basiert, quasi mit verschachtelten if-Statements. Was wir auch hatten ist ein automatisierter Newsletter mit Corona Daten, mit dem Ziel der Bottom-up Darstellung der Daten, um zu sehen, was in der Umgebung passiert. Das wurde mittels automatisiert generierter Grafiken umgesetzt, ohne Textgenerierung.

   Bei [Unternehmen] gab es ein Projekt zur automatischen Wahlberichterstattung, mit der Prämisse, für jeden Wahlkreis einen Text zu generieren. Das wurde bisher mit Templates gelöst, wobei auch mit NLG Libraries experimentiert wurde, um die korrekte Deklination (z.B. Zahl, Geschlecht) der unterschiedlichen Parteien und anderer Wörter zu berücksichtigen. Hierzu haben wir SimpleNLG verwendet, was Java-basiert ist. Hier haben wir relativ schnell gemerkt, dass es im Vergleich zu rein Template-basierten Lösungen deutlich mehr Rechenleistung beansprucht. Vom Ergebnis her war es zum Beispiel schwer Konjunktivsätze zu konstruieren, weil die Modelle [von SimpleNLG] recht einfach gestrickt waren, optimiert auf einfache Lesbarkeit.

   Es ist uns aufgefallen, dass bei der letzten Wahl sehr viele Newsrooms Textgenerierung verwendet haben. Hier hatte ich das Gefühl, dass dies großteils Template-basiert war.

   1.1 [EN] *(1.1-1.4: If technology for automation is in use)*
   Is the software that is in use developed by your company itself or is it licensed from a third-party supplier?

   [DE] *(1.1-1.4: Wenn Technologie zur Automatisierung in Verwendung)*
   Ist die verwendete Software eine Eigenentwicklung Ihres Unternehmens oder wird sie von einem Drittanbieter lizensiert?

   Alles, was wir verwendet haben, war Open-Source und die Software zur Umsetzung war selbst geschrieben. Wir haben also keine Third-party UI-Elemente oder ähnliches verwendet. Ich glaube schon, dass andere Newsrooms mit kommerziellen Anbietern zusammenarbeiten.

1.1.1 [EN] *(If licensed)*

Was it an option to start developing one's own software? Why (not)?

[DE] *(Wenn lizensiert)*

War es eine Option, die Software selbst zu entwickeln? Warum (nicht)?

1.1.2 [EN] *(If self-developed)*

What were the reasons for an in-house development? What benefits and challenges emerged from this?

[DE] *(Wenn Eigenentwicklung)*

Was waren die Gründe für eine Eigenentwicklung? Welche Vorteile und Herausforderungen haben sich dadurch ergeben?

Der große Vorteil für uns war, dass wir alles selbst testen konnten und der ganze Stack bei uns selbst läuft. Wir konnten alleine oder zu zweit den vollen Überblick über jeden Code bewahren, der in der Wahlnacht laufen muss. Das ist auch sehr wichtig bei Projekten, wo es wirklich um etwas geht, wo wenn in zwei Stunden die Sache nicht läuft, man schon später dran ist als alle anderen und keine SEO Benefits mehr hat. Man will auch gegenüber dem Newsroom garantierten, dass wenn man dieses Projekt anstrebt, dieses auch in der Wahlnacht läuft. Da ist es natürlich besser, den Code in der Hand zu haben, sodass man das auch gewährleisten kann und bei Problemen in der Wahlnacht selbst Hand anlegen kann. Wir haben gemerkt, dass die Testdaten nicht ganz genau die Realität der Wahl widergespiegelt haben und haben hier noch in der Wahlnacht Anpassungen vornehmen müssen. Wenn das nicht möglich gewesen wäre, dann wäre das Projekt aufgrund von leicht behebbaren Problemen vielleicht gar nicht erfolgreich gewesen. Ich glaube es hat riesige Vorteile, wenn man die Leute und das Geschehen quasi bei der Hand hat und sich nicht auf Drittanbieter verlassen muss bei solchen Projekten.

1.2 [EN] **From your current perspective, what is your company's biggest challenge when it comes to making one's own software solutions?**

[DE] **Was ist aus heutiger Sicht die größte Herausforderung für Ihr Unternehmen, wenn es um die Entwicklung eigener Softwarelösungen geht?**

Besonders in der Textanalyse ist es nicht so, dass man als Newsroom jeden Tag ein Projekt vor sich sieht, wo man sich denkt, das wäre ein super Anwendungsfall für Textgenerierung. Es gibt ganz wenige Fälle, wo Daten tatsächlich zu dem Zeitpunkt vorliegen, an dem sie auch für die journalistische Bearbeitung relevant sind und wo die journalistische Bearbeitung dann auch noch besonders schnell gehen muss. Da gibt es in Nachrichtenorganisationen beispielsweise zur Grafikproduktion in Wahlnächten etablierte Prozesse und es ist auch immer klar, dass dies so eine journalistische Relevanz hat, dass man auch Zeit in die Vorbereitung stecken will. Die große Problematik bei Eigenprogrammierpro-

91

jekten ist die Vorlaufzeit, die man braucht. Hier passt der Rhythmus oft nicht mit dem restlichen Unternehmen zusammen, besonders mit der Redaktion. Normalerweise muss jeden Tag die Zeitung rausgehen und man muss sich wirklich die Freiheit nehmen und nehmen können, sich aus diesen täglichen schnelllebigen Abläufen herauszunehmen, was in dieser Branche eine gewisse Herausforderung ist. Natürlich setzt das voraus, bereits ein Datenteam in der Redaktion zu haben, das auch Programmieren beherrscht.

1.3 [EN] **What are the biggest benefits that come along with the software solution deployed in your company?**

[DE] **Was sind die größten Vorteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?**

Die Ressourcen, in einer Nacht Texte für alle Wahlkreise händisch zu schreiben, hat niemand. Wir haben uns angesehen, wie oft diese Artikel aufgerufen wurden. Es wurden die meisten Wahlkreise auch aufgerufen und wir hatten insgesamt sehr gute Zugriffszahlen. Das Interesse der Leser besteht also und der Aufwand ist daher gut bei unseren Lesern aufgenommen.

1.4 [EN] **What are the biggest challenges or downsides that come along with the software solution deployed in your company?**

[DE] **Was sind die größten Herausforderungen oder Nachteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?**

Der Nachteil ist, dass alle Informationen von Relevanz in den Daten abgebildet sein müssen, das ist Grundvoraussetzung. Wenn in der Wahlnacht die Daten plötzlich anders ankommen als erwartet, dann führt das zu Problemen, die man schnell lösen muss. Ansonsten ist es sehr abhängig von der Technologie und den Anforderungen, die man an die Technologie stellt. Man kann alles mit Templates machen, muss dann aber noch viel strukturierter testen, um mögliche Fehler auch wirklich zu finden. Ich glaube aber eine Sache, die wir noch nicht so gut gelöst hatten, ist alle möglichen Kombinationen verlässlich durchzutesten. Das ist auf jeden Fall eine Herausforderung, wo man das nächste Mal sehen wird, ob man diese Testproblematik systematischer angehen könnte als wir es diesmal gemacht haben.

2. [EN] *(If no automation tools adopted yet)*
Has the introduction of software automation tools ever been considered in your company so far? If so, which ones?

[DE] *(Wenn bisher noch keine Automatisierungstools in Verwendung)*
Wurde die Einführung von softwarebasierten Automatisierungstools in Ihrem Unternehmen jemals in Betracht gezogen? Wenn ja, welche?

2.1 [EN] **Can you think of workflows in a newsroom that would be well suited to be supported or taken over by intelligent technologies?**

92

[DE] **Fallen Ihnen Workflows in der Redaktion ein, die sich eignen würden, von intelligenten Technologien unterstützt oder ersetzt zu werden?**

3. [EN] *(If no automated news generation adopted yet)*
   In your opinion, what are the key reasons that kept [company name] from adopting automated news generation so far?

   [DE] *(Wenn bisher noch keine automatisierte Generierung von News in Verwendung)*
   Was sind Ihrer Meinung nach die Hauptgründe, die [Unternehmen] bisher davon abgehalten haben, die automatisierte Generierung von News einzuführen?

4. [EN] **In your opinion, how does [company name] currently differentiate itself from competitors?**

   [DE] **Wie differenziert sich [Unternehmen] Ihrer Meinung nach aktuell von Konkurrenten?**

   4.1 [EN] **With an extensive adoption of automated journalism in the future, how might this current key differentiating characteristics be subject to change?**
   [DE] **Mit einer zukünftig umfassenden Einführung von Automated Journalism, wie könnten sich diese aktuellen Hauptdifferenzierungsmerkmale künftig verändern?**

   Ich hege die Hoffnung, dass „Automated Journalism" an Relevanz verliert. Weil Automated Journalism funktioniert nur in Fällen, wo die Daten super schnell da sind und zu diesem Zeitpunkt auch relevant sind. Das ist jetzt mit Corona oder an Wahlabenden der Fall. Wahlabende werden hoffentlich bleiben, aber die ganze Corona Sache wird sich ja hoffentlich beruhigen. Wenn man sich jetzt ansieht, dass Menschen schon bezüglich Ukrainekrieg automatisiert Daten verarbeiten müssen, dann ist das mehr oder weniger ein Thema von dem ich hoffen würde dass es verschwindet und dass die journalistische Coverage auch auf Themen geht, die nicht die nächste große Krise sind. In Zukunft nehme ich an, dass Klimaberichterstattung auch mehr aufgrund von Daten geschehen wird, beispielsweise zu lokalen Auswirkungen. Also insgesamt ist meine Hoffnung, dass es weniger an Relevanz gewinnt und man sich nicht durch Textautomatisierung abheben muss, sondern durch klassischen, guten Journalismus.

Interview C

1. [EN] **Does [company name] currently use intelligent technologies for the automation of workflows, particularly automated news generation? If so, which ones?**

   [DE] **Nutzt [Unternehmen] aktuell intelligente Technologien für die Automation von Workflows, insbesondere zur automatischen Generierung von News? Wenn ja, welche?**

   Nein, keine automatische Newsgenerierung. Was wir in der Redaktion aber probieren, ist die Anfertigung von Interviewtranskripten durch Software. Das machen wir zum Teil über Adobe Premiere Pro. Ich habe für mich aus persönlichem Interesse alle bekannten Plattformen durchgetestet - da gibt es Plattformen wie „Trint" oder „Otter", was die größten Plattformen für Interviewtranskription sind. Diese kosten jedoch alle ziemlich viel, verlangen eine wirklich schöne Summe monatlich, sind aber für den englischen Raum getrimmt. Sprich, wenn man es dann auf Deutsch austestet, nimmt es einem logischerweise einen Teil der Arbeit ab, weil ein Teil des Transkript ist halbwegs ok, aber in der Tiefe noch nicht. Wenn ich mir dann also trotzdem wieder das ganze Interview anhören muss und jedes vierte Wort füllen muss, dann kann ich es am Ende gleich selbst machen und mir das Geld sparen. Die Qualität stimmt hier also noch nicht. Sollten Sie über eine stolpern, die in den Hundert, die ich getestet habe, noch nicht dabei ist, dann geben Sie mir bitte Bescheid. Wenn es eine Software gäbe, die mir wirklich ein Interview eins zu eins transkribiert, dann wäre unsere Redaktion glaube ich sofort bereit auch viel Geld dafür auszugeben. Das würde nämlich sehr viele Ressourcen freimachen. Nur da sind halt alle die momentan am Markt sind noch nicht weit genug. Für mich ein Paradebeispiel sind Smarthome-Assistants/Sprachassistenten. Wenn für diese Teile, wo unfassbar viel Entwicklungsgeld hineinfließt, wenn sogar die, übertrieben gesagt, nur jeden vierten Satz richtig verstehen, wie soll dann die KI ein Interview perfekt transkribieren können. Da fehlt offenbar noch ein wenig. Was bei uns noch dazukommt ist der Nachteil des Österreichers, denn Dialekte helfen bei so etwas logischerweise gar nicht. Das ist mir schon aufgefallen, wenn ich Interviews mit einem bundesdeutschen Kollegen führe. Dann sind die Transkripte besser als mit einem Österreicher.

   1.1 [EN] *(1.1-1.4: If technology for automation is in use)*
       Is the software that is in use developed by your company itself or is it licensed from a third-party supplier?

       [DE] *(1.1-1.4: Wenn Technologie zur Automatisierung in Verwendung)*
       Ist die verwendete Software eine Eigenentwicklung Ihres Unternehmens oder wird sie von einem Drittanbieter lizensiert?

       1.1.1 [EN] *(If licensed)*
             Was it an option to start developing one's own software? Why (not)?

94

[DE] *(Wenn lizensiert)*

War es eine Option, die Software selbst zu entwickeln? Warum (nicht)?

1.1.2 [EN] *(If self-developed)*

What were the reasons for an in-house development? What benefits and challenges emerged from this?

[DE] *(Wenn Eigenentwicklung)*

Was waren die Gründe für eine Eigenentwicklung? Welche Vorteile und Herausforderungen haben sich dadurch ergeben?

1.2 [EN] **From your current perspective, what is your company's biggest challenge when it comes to making one's own software solutions?**

[DE] **Was ist aus heutiger Sicht die größte Herausforderung für Ihr Unternehmen, wenn es um die Entwicklung eigener Softwarelösungen geht?**

1.3 [EN] **What are the biggest benefits that come along with the software solution deployed in your company?**

[DE] **Was sind die größten Vorteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?**

1.4 [EN] **What are the biggest challenges or downsides that come along with the software solution deployed in your company?**

[DE] **Was sind die größten Herausforderungen oder Nachteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?**

2. [EN] *(If no automation tools adopted yet)*
Has the introduction of software automation tools ever been considered in your company so far? If so, which ones?

[DE] *(Wenn bisher noch keine Automatisierungstools in Verwendung)*
Wurde die Einführung von softwarebasierten Automatisierungstools in Ihrem Unternehmen jemals in Betracht gezogen? Wenn ja, welche?

Ja und nein. Es ist so, dass wir mit [Unternehmen] Gespräche hatten. [Unternehmen] hat hier einen ersten Textbot entwickelt den sie uns vorgestellt haben. [Unternehmen] hat also solche Technologien in Aussicht gestellt. Da ist meines Wissens nach aber der nächste Schritt, sprich die wirkliche Massentauglichkeit, sodass es im Aktionsbetrieb gut funktionieren kann, noch nicht gemacht worden und deshalb noch nicht im Einsatz.

2.1 [EN] **Can you think of workflows in a newsroom that would be well suited to be supported or taken over by intelligent technologies?**

[DE] **Fallen Ihnen Workflows in der Redaktion ein, die sich eignen würden, von intelligenten Technologien unterstützt oder ersetzt zu werden?**

Natürlich. Ich muss sagen ich habe hier weniger Berührungsängste als wahrscheinlich viele Kollegen und habe auch selbst bereits mehrere Geschichten darüber geschrieben. Ich glaube, das was den Journalisten reizt, ist die gute Geschichte. Und die gute Geschichte, die tiefe Geschichte, die investigative Geschichte, lebt von zweierlei Ebenen. Das ist zum einen die Recherche und hier rede ich nicht davon, dass ich Wikipedia Artikel zusammenkopiere, sondern an die richtigen Interviewpartner zu kommen, diese auch bewusst auszusuchen, mit diesen Gespräche führen, oder auch investigativ tätig sein, Exklusivinfos aufstellen, etc., etc. Das wird einem eine KI nicht abnehmen können. Zum anderen das tolle Storytelling. Damit meine ich Geschichten wirklich bunt schreiben und wirklich Kopfkino entstehen lassen. Und auch da tut sich meines Wissens die KI, Stand heute, noch sehr schwer. Das heißt für diese zwei Ebenen, die den Journalisten am meisten reizen, da sehe ich gerade wenig Alternativen.

Was aber die KI gut kann, ist simple Textbausteine abzunehmen. Sprich Börsenmeldungen, Wettermeldungen, von mir aus sogar eine Agenturmeldung einer Presseagentur, von den 20 ausgeschickten Zeilen zusammengekürzt auf 5 Zeilen, für unsere Kurzmeldungen. Solche Anwendungen finde ich sinnvoll, allein schon, weil es journalistische Ressourcen freispielt. Also wenn die KI das machen kann, dann kann der Journalist sich auf das konzentrieren, wofür er eigentlich da ist, und das was er auch wirklich kann. Das ist eben dieser Investigativhintergrund/ die Storytelling Schiene.

3. [EN] *(If no automated news generation adopted yet)*
   In your opinion, what are the key reasons that kept [company name] from adopting automated news generation so far?

   [DE] *(Wenn bisher noch keine automatisierte Generierung von News in Verwendung)*
   Was sind Ihrer Meinung nach die Hauptgründe, die [Unternehmen] bisher davon abgehalten haben, die automatisierte Generierung von News einzuführen?

   Das ist eine Entscheidung der Geschäftsführung/Redaktionsleitung, also weniger eine welche die Redakteure selbst treffen. Da ist es meines Wissens nach eine Angebot/Nachfrage Geschichte. Also wir haben es jetzt sicher nicht händeringend gesucht, aber wenn ich jetzt nicht völlig falsch liege, es gibt am Markt jetzt nicht so wahnsinnig viel an guter Software, vor allem für den deutschsprachigen Raum. Im Englischsprachigen sind wir glaube ich noch ein Stück weiter und ich glaube das bekannteste Beispiel ist ja dieser Erdbeben-Bot, womit automatische Berichterstattung zu Erdbeben betrieben wird. Das geht im Englischen aber auch viel besser, weil diese KIs im Regelfall auf die englische Sprache trainiert sind; vielleicht gibt es da im Asiatischen/Chinesischen auch etwas. Im Deutschen sind die KIs eben noch eine Spur schwächer. Also meines Wissens gibt es noch nicht wahnsinnig viel für den deutschsprachigen Raum und wir haben es auch nicht händeringend gesucht. Also die Summe daraus ergibt, dass wir es [automatisch generierte News] noch nicht

haben. Was man auch sagen muss, am Ende ist es auch eine Kosten/Nutzen Frage. Wenn ich weiß ich schaufle am Tag angenommen 4 Stunden an Ressourcen für Journalisten frei, weil eine KI irgendwelche Kurzmeldungen abnimmt, diese KI für diese 4 Stunden aber mehr als ein angestellter Journalist kostet, dann würde das nicht für die KI sprechen. Kosten/Nutzen muss hier also wie immer auch betrachtet werden.

4. [EN] **In your opinion, how does [company name] currently differentiate itself from competitors?**

[DE] **Wie differenziert sich [Unternehmen] Ihrer Meinung nach aktuell von Konkurrenten?**

Ich denke der österreichische Zeitungsmarkt ist ein sehr kleiner, ein sehr überschaubarer, im Vergleich zu anderen Ländern. Wir haben 13 Tageszeitungen, nageln sie mich nicht auf die genaue Zahl fest, aber ich glaube die Schweiz hat 70, Schweden hat 120, teilweise noch mit Tages- und Abendausgaben. Von so etwas sind wir weit weg. Weil wir eben einen so kleinen Markt haben, ist es ein bisschen leichter generalistisch unterwegs zu sein, ohne sich ganz stark zu spezialisieren. Trotzdem funktioniert es auch in Österreich eigentlich nur wenn du für etwas stehst. [Unternehmen des Interviewpartners] hat hier im Grunde zweierlei Ebenen. Zum einen sind wir einer der, ich sage jetzt einmal ganz arrogant, drei größten richtigen Qualitätszeitungen Österreichs, differenzieren uns zum einen also durch den Qualitätsanspruch. Zum anderen, auch wenn wir national erscheinen und daher auch viele Leser außerhalb des Bundeslandes haben, haben wir einen ganz starken Fokus für unsere Region. Wenn man also zu uns greift, hat man garantiert Nachrichten aus der Region. Wenn wir noch ins Detail gehen, dann gibt es bestimmte Eigenschaften des Hauses, gewachsen über die Jahre, die uns anders oder besser machen als die Konkurrenz. Wir sind zum Beispiel kulturell und auch innenpolitisch sehr stark aufgestellt.

4.1 [EN] **With an extensive adoption of automated journalism in the future, how might this current key differentiating characteristics be subject to change?**

[DE] **Mit einer zukünftig umfassenden Einführung von Automated Journalism, wie könnten sich diese aktuellen Hauptdifferenzierungsmerkmale künftig verändern?**

Das ist eine gute Frage. Ich sehe KI und auch Automated Journalism uns nur für Kleinarbeiten oder Zuarbeiten helfend, also eben für Interviewtranskripte oder für Kurzmeldungen. Wenn die KI das wirklich zu einem machbaren/leistbaren Preis bieten kann, dann spült das für uns eben genau Sachen frei für die wir stehen, nämlich Qualitätsjournalismus. Das lässt uns das was ich ihnen vorhin aufgezählt habe noch breiter und besser machen, nämlich noch bessere Regionalberichterstattung, noch bessere Innenpolitikberichterstattung und noch bessere Kulturberichterstattung.

Interview D

1. [EN] **Does [company name] currently use intelligent technologies for the automation of workflows, particularly automated news generation? If so, which ones?**

   [DE] **Nutzt [Unternehmen] aktuell intelligente Technologien für die Automation von Workflows, insbesondere zur automatischen Generierung von News? Wenn ja, welche?**

   Was wir bisher gemacht haben ist die automatisierte Erstellung von Texten in der Berichterstattung über Wahlergebnisse, für Wahlen auf Bundes- und Landesebene. Das läuft auch weiterhin noch. Das ist der einzige Case der mir einfällt, der live gegangen ist. Wir haben es anhand von anderen Usecases mehrfach überprüft, also beispielsweise Sport- oder Börsenberichterstattung, aber diese Anläufe haben dann aus unterschiedlichen Gründen nicht funktioniert. Aber das jetzige Verfahren ist bei der Wahl auf Bundesebene zum ersten Mal live gewesen und das machen wir auch bei Landtagswahlen.

   Was wir uns auch angeschaut haben, ist die automatische Übersetzung von Texten in andere Sprachen, um neue Märkte stärker zu erschließen. Das könnte schon sein, dass wir solche Modelle auf Wiedervorlage nehmen und uns immer mal wieder anschauen, haben es jetzt provisorisch aber erstmal verworfen in der letzten Zeit. Aber wenn die Übersetzungstechnologien da noch besser werden, dann könnte das in den nächsten Jahren anders aussehen. Ich glaube vor Kurzem hat „Le Monde" ein englisches Angebot gestartet, das glaube ich zumindest auf automatischer Übersetzung fußt. Solche Modelle haben wir auch angeschaut, aber erstmal verworfen, weil bei der Art von Texten, die wir haben, die auch häufig justiziabel sind, eine reine automatisierte Übersetzung der Texte, ohne anschließender Prüfung durch die Redaktion, ob das auch alles richtig und rechtsfest ist, uns zu riskant erschien und dadurch erstmal der Business Case in sich zusammengefallen ist.

   1.1 [EN] *(1.1-1.4: If technology for automation is in use)*
   Is the software that is in use developed by your company itself or is it licensed from a third-party supplier?

   [DE] *(1.1-1.4: Wenn Technologie zur Automatisierung in Verwendung)*
   Ist die verwendete Software eine Eigenentwicklung Ihres Unternehmens oder wird sie von einem Drittanbieter lizensiert?

   Das ist ein Drittanbieter, über die Firma Retresco, die uns da ein Framework anbieten, das wir noch bestücken und mit Regeln versehen. Das haben wir aber im Kern nicht selbst entwickelt.

   1.1.1 [EN] *(If licensed)*
   Was it an option to start developing one's own software? Why (not)?

[DE] *(Wenn lizensiert)*

War es eine Option, die Software selbst zu entwickeln? Warum (nicht)?

Tatsächlich ja, denn so wahnsinnig kompliziert sind die Regeln, die dahinter liegen, auch wieder nicht. Aber dadurch, dass das ganze Userinterface bei Retresco schon dasteht, hat uns das dann überzeugt, dass wir dachten, da sind bestimmte Dinge, die wir nicht erledigen müssen, sodass wir da jetzt tatsächlich nur die Regeln eintragen. So ist das Projekt für uns handhabbarer geworden.

1.1.2 [EN] *(If self-developed)*

What were the reasons for an in-house development? What benefits and challenges emerged from this?

[DE] *(Wenn Eigenentwicklung)*

Was waren die Gründe für eine Eigenentwicklung? Welche Vorteile und Herausforderungen haben sich dadurch ergeben?

1.2 [EN] From your current perspective, what is your company's biggest challenge when it comes to making one's own software solutions?

[DE] **Was ist aus heutiger Sicht die größte Herausforderung für Ihr Unternehmen, wenn es um die Entwicklung eigener Softwarelösungen geht?**

Es ist schon so, dass das nicht Kernkompetenz unserer Entwickler ist. Also IT-Kapazität haben wir relativ viel, die ist aber auf andere Bereiche konzentriert, sodass es für uns hier ökonomisch sinnvoller ist da zu partnern, statt eigene Entwicklerressourcen aufzuwenden. Wir konzentrieren uns dann auf das Erstellen der Regeln. Das ist eine Aufgabe, die bei uns von Journalisten gemacht wird. Es ist für uns so also möglich, unsere IT da rauszuhalten. Insofern ist die größte Herausforderung bei uns die Verfügbarkeit von IT-Ressourcen, oder eine andere Priorisierung von IT-Ressourcen.

1.3 [EN] What are the biggest benefits that come along with the software solution deployed in your company?

[DE] **Was sind die größten Vorteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?**

Die Software ist mit ein bisschen Affinität und Programmierverständnis relativ leicht zu bedienen. Also der hohe Komfort in der Bedienbarkeit ist auf jeden Fall ein Plus.

1.4 [EN] What are the biggest challenges or downsides that come along with the software solution deployed in your company?

[DE] **Was sind die größten Herausforderungen oder Nachteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?**

Der Aufwand ist trotzdem nicht zu vernachlässigen, sodass wir uns wenn dann Usecases aussuchen welche eine hohe Reproduzierbarkeit haben, sprich

Landtagswahlen. Man kann die Regeln, die man da schreibt, weitestgehend wiederholen und bei den anderen Beispielen, an die wir gedacht haben, wäre das auch so. Das Thema, das man sich aussucht, braucht eine hohe Konstanz, etwas das regelmäßig wiederkehrt.

2. [EN] *(If no automation tools adopted yet)*
   Has the introduction of software automation tools ever been considered in your company so far? If so, which ones?

   [DE] *(Wenn bisher noch keine Automatisierungstools in Verwendung)*
   Wurde die Einführung von softwarebasierten Automatisierungstools in Ihrem Unternehmen jemals in Betracht gezogen? Wenn ja, welche?

   2.1 [EN] **Can you think of workflows in a newsroom that would be well suited to be supported or taken over by intelligent technologies?**
   [DE] **Fallen Ihnen Workflows in der Redaktion ein, die sich eignen würden, von intelligenten Technologien unterstützt oder ersetzt zu werden?**

3. [EN] *(If no automated news generation adopted yet)*
   In your opinion, what are the key reasons that kept [company name] from adopting automated news generation so far?

   [DE] *(Wenn bisher noch keine automatisierte Generierung von News in Verwendung)*
   Was sind Ihrer Meinung nach die Hauptgründe, die [Unternehmen] bisher davon abgehalten haben, die automatisierte Generierung von News einzuführen?

4. [EN] **In your opinion, how does [company name] currently differentiate itself from competitors?**

   [DE] **Wie differenziert sich [Unternehmen] Ihrer Meinung nach aktuell von Konkurrenten?**

   Ich komme mal von dem Automatisierungsaspekt. Ich glaube wir sind da noch relativ defensiv. Ich glaube bestimmte andere Unternehmen sind da schon deutlich stärker am Experimentieren. Das sind teilweise lokale Anbieter. Das liegt daran, dass für eine Lokalzeitung bestimmte Informationen wie Verkehrsnachrichten oder Wetter zum Markenkern gehören können, während [Unternehmen des Interviewpartners], als ein politisches investigativ geprägtes Medium, solche Themenfelder nicht natürlicherweise in seinem Markenkosmos hat. Deswegen haben auch Felder wie Börsenberichterstattung, was ja auch Wirtschaftsmedien teilweise machen, oder Sportautomatisierung, zu uns nicht so sehr gepasst. Daher haben wir uns dann am Ende für Politik entschieden. Das ist die Themenausrichtung unserer Marke. Wir haben in den letzten Jahren auch eine stärkere Orientierung von einer Werbeorientierung im Digitalen hin zu einem Abonnement-System. Das Geschäftsmodell ist also anders, sodass die reine Ausrichtung auf vermarktbare Page-Impressions

100

für uns gar nicht mehr so wichtig ist. Automatisierter Journalismus ist natürlich vor allem interessant, wenn man eine Masse an Klicks oder unique Users machen möchte. Für uns sind aber Nutzer besonders dann interessant, wenn wir sie perspektivisch irgendwann einmal in ein Abonnement überführen können. Insofern hat automatisierter Journalism da für uns eine geringere Priorität als für bestimmte andere Titel.

4.1 [EN] **With an extensive adoption of automated journalism in the future, how might this current key differentiating characteristics be subject to change?**

[DE] **Mit einer zukünftig umfassenden Einführung von Automated Journalism, wie könnten sich diese aktuellen Hauptdifferenzierungsmerkmale künftig verändern?**

Wenn Automated Journalism bei der Konkurrenz stärker eingeführt wird, glaube ich würde sich für uns erstmal nicht unbedingt etwas ändern. Es sei denn es findet sich ein neuer Weg der Monetarisierung im Rahmen unseres Abonnement-Modells, den ich aber im Moment nicht sehen würde.

101

Interview E

1. [EN] **Does [company name] currently use intelligent technologies for the automation of workflows, particularly automated news generation? If so, which ones?**

   [DE] **Nutzt [Unternehmen] aktuell intelligente Technologien für die Automation von Workflows, insbesondere zur automatischen Generierung von News? Wenn ja, welche?**

   Also da hätten wir mal das Fußballprojekt. Da decken wir sehr viele Ligen ab, Unterligen, also Amateurfußball. Zusätzlich haben wir diverse Projekte, wo wir mit automatisierten Verfahren News produzieren - das ist quasi ein Projekt, das mehrere Unterprojekte hat, beispielsweise individuelle News auf Gemeindeebene.

   Alles was bisher produziert wurde ist Template-basiert, weil wir damit einfach unter Kontrolle haben, was dabei herauskommt. Wir beobachten die Szene mit Machine Learning und Deep Learning und was es da noch alles gibt, sind aber der Meinung, dass das noch zu wenig ausgereift ist, um im journalistischen Kontext eingesetzt zu werden.

   1.1 [EN] *(1.1-1.4: If technology for automation is in use)*
   Is the software that is in use developed by your company itself or is it licensed from a third-party supplier?

   [DE] *(1.1-1.4: Wenn Technologie zur Automatisierung in Verwendung)*
   Ist die verwendete Software eine Eigenentwicklung Ihres Unternehmens oder wird sie von einem Drittanbieter lizensiert?

   Wir nutzen hier eine Drittanbieterlösung, die auch bei anderen Nachrichtenorganisationen zum Einsatz kommt. Das ist eine Templating-Plattform die wir nutzen, um die Textvorlagen zu schreiben. Früher bei [Unternehmen] hatten wir eine Eigenentwicklung.

   1.1.1 [EN] *(If licensed)*
   Was it an option to start developing one's own software? Why (not)?
   [DE] *(Wenn lizensiert)*
   War es eine Option, die Software selbst zu entwickeln? Warum (nicht)?
   Es war schon auch eine Option das selbst zu entwickeln. Wir haben uns dann für eine fertige Softwarelösung entschieden, einfach um das auch einmal auszuprobieren. Wir wollten möglichst schnell zu Resultaten kommen und da ist das dann halt einfach schneller und rascher, wenn man eine fertige Software verwendet. Das ist eigentlich der Hauptgrund. Auch meine Erfahrung mit der Eigenentwicklung hat gezeigt, dass es sehr aufwendig ist das zu warten. Von daher ist es sehr komfortabel eine Fertiglösung zu haben.

102

1.1.2 [EN] *(If self-developed)*

What were the reasons for an in-house development? What benefits and challenges emerged from this?

[DE] *(Wenn Eigenentwicklung)*

Was waren die Gründe für eine Eigenentwicklung? Welche Vorteile und Herausforderungen haben sich dadurch ergeben?

Damals [bei anderem Unternehmen] wollten wir auch rasch zu einer Lösung gelangen und deshalb wurde dann ein sehr spezialisiertes Tool entwickelt, welches das Thema Abstimmungen beherrscht, aber für alle anderen Themen wäre es dann wieder schwierig gewesen das umzubauen. Für diesen spezialisierten Fall war eine Eigenentwicklung besser. Wenn man es dann breiter einsetzen will, zum Beispiel für etwas so Komplexes wie die Fußballberichterstattung, wollten wir uns auf die Textvorlagen konzentrieren und nicht auf die Software.

1.2 [EN] **From your current perspective, what is your company's biggest challenge when it comes to making one's own software solutions?**

[DE] **Was ist aus heutiger Sicht die größte Herausforderung für Ihr Unternehmen, wenn es um die Entwicklung eigener Softwarelösungen geht?**

Eine eigene Software ist halt ein Commitment zu einem Produkt, das man dann auch weiterentwickeln, warten und laufend an Veränderungen im gesamten System anpassen muss. Ich glaube sowas müsste man dann ins Auge fassen, wenn man genau weiß, was man langfristig benötigt. Der Vorteil einer Software, die man einkauft, ist, dass man schnell Resultate bekommt. Es kann sein, dass man irgendwann den Fokus ändert und doch etwas selbst bauen will, aber man geht dann eine große Verpflichtung ein. In unserem speziellen Fall hätten wir auch gar nicht die Entwickler dazu gehabt, einfach unternehmensbedingt.

1.3 [EN] **What are the biggest benefits that come along with the software solution deployed in your company?**

[DE] **Was sind die größten Vorteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?**

Wir kommen damit in die Lage, dass wir hyperlokale Inhalte bieten können, die wir sonst nicht bieten könnten. Wir können unseren Lesern damit einen Zugang anbieten, der sehr auf ihre Lebenswelt zugeschnitten ist. Vor allem für ihre Gemeinde, das ist bei uns die kleinste politische Einheit. Wir sind in der Lage Texte zu generieren, die sich genau mit einer Gemeinde befassen. Oder sei es mit dem Fußballclub in deren Gemeinde. Also diese Kleinteiligkeit, diese Granularität, ist ganz sicher der größte Vorteil der automatisierten Berichterstattung.

1.4 [EN] **What are the biggest challenges or downsides that come along with the software solution deployed in your company?**

Boilerplate side text on left margin

[DE] **Was sind die größten Herausforderungen oder Nachteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?**

Die große Herausforderung ist, dass diese Inhalte irgendwie an die Leser und Leserinnen gelangen müssen. Man kann beispielsweise nicht einfach 700 Texte auf die Frontseite spielen und annehmen, dass diese von den Lesern selbst gefunden werden. Man muss „Gefäße" finden, wie man den Lesern diese Fülle überhaupt anbieten kann. Das ist eine große Herausforderung.

Nachteilig ist, dass diese automatisierten Inhalte sehr faktenbasiert sind und vom Storytelling her keine extrem spannenden Texte sind. Man darf es da halt auch nicht übertreiben, also muss das richtige Maß finden, weil man dem Leser und der Leserin ja auch nicht nur solche Texte bieten will, sondern eben auch solche mit Menschen involviert, wo Recherche dahintersteckt und wo auch mit Protagonisten gesprochen wurde.

2. [EN] *(If no automation tools adopted yet)*
   Has the introduction of software automation tools ever been considered in your company so far? If so, which ones?

   [DE] *(Wenn bisher noch keine Automatisierungstools in Verwendung)*
   Wurde die Einführung von softwarebasierten Automatisierungstools in Ihrem Unternehmen jemals in Betracht gezogen? Wenn ja, welche?

   2.1 [EN] **Can you think of workflows in a newsroom that would be well suited to be supported or taken over by intelligent technologies?**
   [DE] **Fallen Ihnen Workflows in der Redaktion ein, die sich eignen würden, von intelligenten Technologien unterstützt oder ersetzt zu werden?**

3. [EN] *(If no automated news generation adopted yet)*
   In your opinion, what are the key reasons that kept [company name] from adopting automated news generation so far?

   [DE] *(Wenn bisher noch keine automatisierte Generierung von News in Verwendung)*
   Was sind Ihrer Meinung nach die Hauptgründe, die [Unternehmen] bisher davon abgehalten haben, die automatisierte Generierung von News einzuführen?

4. [EN] **In your opinion, how does [company name] currently differentiate itself from competitors?**

   [DE] **Wie differenziert sich [Unternehmen] Ihrer Meinung nach aktuell von Konkurrenten?**

   Wir sind in einem relativ großen Teil des Landes aktiv, in vielen unterschiedlich großen Gemeinden. Das ist das ideale Spielfeld, um mit automatisierten Meldungen für all diese Gemeinden irgendetwas zu bieten. Von daher sind wir schon echt

gut prädestiniert für diese Automatisierungslösungen und das setzt uns zu einem gewissen Teil auch von unserer Konkurrenz ab, die viel mehr als wir auf einzelne große Publikationen ausgerichtet sind, sei es eine große Boulevardzeitung oder sei es eine national ausstrahlende Zeitung, in anderen Größen als wir das haben. Also wir sind sehr regional ausgerichtet und das hilft bei der Automatisierung auch.

4.1 **[EN] With an extensive adoption of automated journalism in the future, how might this current key differentiating characteristics be subject to change?**

[DE] Mit einer zukünftig umfassenden Einführung von Automated Journalism, wie könnten sich diese aktuellen Hauptdifferenzierungsmerkmale künftig verändern?

Ich gehe davon aus, dass alle Medien früher oder später in diesem Feld aktiv werden. Gerade Fußball ist ein Thema, das es überall gibt und da kann dieser Regionalisierungsansatz, der bei uns jetzt sehr stark ist, irgendwann ein bisschen in den Hintergrund rücken – also dass sich andere Medien da auch quasi in unserem Feld austoben könnten. Weil der Eintrittspreis ist relativ gering – man muss in diese Automatisierungslösung investieren, aber wenn das einmal getan ist, dann funktioniert das auch. Es ist auch durchaus möglich, dass ein großer Player auf internationaler Ebene in dieses Feld einsteigt. Solche Anzeichen gibt es momentan nicht, aber das kann durchaus sein. Ich denke für ein regionales Medium ist es fast nötig, sich diese Initiativen zumindest einmal angeschaut zu haben. Weil das Wissen dann im Unternehmen zu haben, hilft einem auch das potenzielle Erstarken der Konkurrenz von vornherein zu antizipieren und so darauf reagieren zu können. Schlussendlich muss jedes Unternehmen selbst herausfinden, was eine Automatisierungslösung für dieses Unternehmen auch bieten kann – und ich glaube da gibt es noch ganz viele Nischen.

Interview F

1. [EN] **Does [company name] currently use intelligent technologies for the automation of workflows, particularly automated news generation? If so, which ones?**

   [DE] **Nutzt [Unternehmen] aktuell intelligente Technologien für die Automation von Workflows, insbesondere zur automatischen Generierung von News? Wenn ja, welche?**

   Die automatisierte Wahlberichterstattung ist ein Dauerprojekt. Das ist ein Produkt, das wir schon lange am Laufen haben. Dann haben wir einen Bot, der öffentlich über Social Media zugänglich ist und auch ein bisschen als Showcase dient. Wir verwenden Textautomatisierung für interne Zwecke im Falle der täglichen Corona Berichterstattung, was Infektions-, Test- und Impfgeschehen betrifft. Da bekommt die Redaktion einen Vorschlag für eine Meldung, keine fertige Meldung. Aber die Redaktion kann sich sicher sein, dass alle Zahlen und Interpretationen stimmen, weil das macht die Maschine und dann schreiben die Kollegen und Kolleginnen den Text, der die für den Tag relevanten Aspekte hervorhebt. Eine Anwendung, die bei uns jetzt relativ neu ist, ist eine Ergänzung eines redaktionellen Produktes, bezüglich Branchen- und Wirtschaftsjournalismus. Hier wird die die Insolvenzentwicklung der jeweiligen Branche durch eine Automatisierungskomponente ergänzt.

   1.1 [EN] *(1.1-1.4: If technology for automation is in use)*
   Is the software that is in use developed by your company itself or is it licensed from a third-party supplier?

   [DE] *(1.1-1.4: Wenn Technologie zur Automatisierung in Verwendung)*
   Ist die verwendete Software eine Eigenentwicklung Ihres Unternehmens oder wird sie von einem Drittanbieter lizensiert?

   Da verwenden wir die Software eines Drittanbieters. Es gibt diverse Anbieter am Markt und wir haben da einen aus dem deutschsprachigen Raum genommen. Das hatte damals den Sinn, obwohl auch nicht deutschsprachige Anbieter natürlich Deutsch unterstützen, dass wir uns trotzdem sicherer damit gefühlt haben einen Anbieter zu nehmen, der sozusagen als Erstsprache Deutsch hat. Und das hat sich auch bewährt, weil die trickreiche Falle bei diesem Verfahren ist es, das Sprachmodell für das journalistische Erzählen zum Einsatz zu bringen. Hier hat gerade das Deutsche sehr viele Besonderheiten.

   1.1.1 [EN] *(If licensed)*
   Was it an option to start developing one's own software? Why (not)?
   [DE] *(Wenn lizensiert)*
   War es eine Option, die Software selbst zu entwickeln? Warum (nicht)?

106

Also wir hatten uns damals diese grundsätzliche Frage „make or buy" gestellt, allerdings ist das Entwickeln einer NLG Software nicht unbedingt Kernaufgabe unseres Unternehmens.

1.1.2 [EN] *(If self-developed)*
What were the reasons for an in-house development? What benefits and challenges emerged from this?
[DE] *(Wenn Eigenentwicklung)*
Was waren die Gründe für eine Eigenentwicklung? Welche Vorteile und Herausforderungen haben sich dadurch ergeben?

1.2 [EN] **From your current perspective, what is your company's biggest challenge when it comes to making one's own software solutions?**

[DE] **Was ist aus heutiger Sicht die größte Herausforderung für Ihr Unternehmen, wenn es um die Entwicklung eigener Softwarelösungen geht?**

Wir haben eine sehr große Software Development Abteilung, die hier sehr stark tätig ist, allerdings immer fokussiert auf das Kerngeschäft unseres Unternehmens. Die Anbieter von NLG Software kommen überwiegend aus dem E-Commerce. Das gehört einfach nicht zum Kerngeschäft von uns.

1.3 [EN] **What are the biggest benefits that come along with the software solution deployed in your company?**

[DE] **Was sind die größten Vorteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?**

Die Lösung, die wir im Einsatz haben, und ich kenne andere Lösungen natürlich bei Weitem nicht so gut und habe damit noch nicht gearbeitet, ist für uns deswegen optimal, weil sie noch sehr viel Arbeit an und mit den Daten erlaubt. Also der Content, den man in diesem Verfahren erstellt, ist immer nur so gut wie das Datenmaterial. Im Gegensatz zu E-Commerce Usecases, wo es eigentlich um ein statisches Beschreiben von zum Beispiel Produkten wie Jeans geht, das ist relativ einfach und eigentlich keine Story. Im Journalismus erzählen wir ja Geschichten und diese Geschichten sind oft zu kontextualisieren. Die Informationen in Geschichten beziehen sich auf die Daten, aber müssen schon eine Dateninterpretation haben, damit man diese Geschichte dann weitererzählen kann, also nach A kommt B, und deswegen kommt C, eine Art Kausalität. Das habe ich alles nicht, wenn ich nur ein Produkt beschreibe. Dafür ist es sehr wichtig, dass wir auch in der Software an den und mit den Daten arbeiten können, da auch ein ganzes Transform Modul haben, wo man diverse Mappings machen kann und ein bisschen herumcoden kann, damit wir aus den Daten das beste herausziehen können. Das ist ein entscheidender Vorteil. Eine Lösung wo man sagt, die Daten müssen fix fertig hineinkommen und dann kann man nur noch die Texte konfigurieren, das wäre für uns, mit unserem dezidiert datenjournalistischen Ansatz, nicht zielführend und auch nicht möglich. Auch ein Vorteil ist das Sprachwissen, das diese Software

bereits fertig aufweist. Ich muss Linguistikfunktionen beispielsweise nicht selbst programmieren und kann mich ganz auf den Text konzentrieren und kann mir sicher sein, dass Kasus und Numerus stimmen.

Eine potenzielle zeitliche Entlastung des Personals ist immer abhängig vom Stadium des Projekts. Diese Fantasien, die vielleicht manche haben, dass ich so eine Maschine einsetze und mir dadurch drei Leute erspare, die sind einfach nicht realistisch, weil bei solchen Projekten vor allem der Initialaufwand sehr hoch ist. Das ist eine oft mühsame Kleinarbeit. Man muss sehr viel an den Daten und Texten arbeiten. Wenn das dann einmal steht, dann beginnt sozusagen der Skalierungseffekt schlagend zu werden. Je länger man ein solches System zur Content-Produktion im Einsatz haben kann, desto eher verspürt man dann die Erleichterung. Das Beispiel wäre zum Beispiel unser Corona Textassistent. Jetzt wo das Ding läuft, es läuft ja schon länger, haben wir damit sicher schätzungsweise eine halbe Stunde bis Stunde pro Tag an Arbeit gespart, weil von Dateneingang, der früher manuell via Excel erfolgte, bis zur Produktion des Textvorschlages, wird mittlerweile fast alles vollautomatisiert gemacht und damit ersparen wir uns täglich Zeit. Diese Zeit können wir für unseren eigentlich Job nutzen, nämlich die eigentliche Geschichte zu erzählen. Es ist ja nicht die Kernaufgabe von Journalisten, in Excel Tabellen Zahlen einzutragen oder herauszukopieren. Der Initialaufwand dafür war aber natürlich auch enorm. Hätten wir nach einem Monat gesagt, das war es, das brauchen wir jetzt nicht mehr, dann hätte es sich wahrscheinlich nicht ausgezahlt. Je länger das Ding läuft, desto sinnvoller kann man dann auch tatsächlich Erleichterungen im Arbeitsalltag abgreifen.

1.4 [EN] **What are the biggest challenges or downsides that come along with the software solution deployed in your company?**

[DE] **Was sind die größten Herausforderungen oder Nachteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?**

Der Nachteil ist bis zu einem gewissen Grad aber auch das Sprachwissen, das praktisch nie komplett genug sein kann, um Geschichten so zu schreiben, wie sie ein Mensch erzählen würde. Selbst wenn das Sprachwissen sehr gut ist, ist das nur ein Wissen über die Sprache und nicht über das was dieses System beschreibt, nämlich die Welt. Ich finde es sehr spannend, wie vereinfacht text- oder sprachproduzierende KI-Modelle oft als etwas bewusst Erzählendes dargestellt werden, das ist es nämlich nicht. Gute Modelle wissen, wie die Sprache funktioniert, auch mit Einschränkungen, wissen aber nicht, was diese Sprache beschreibt. Das ist natürlich insofern ein Nachteil, als dass man dann oft mit sehr vielen Tricks versuchen muss, die richtige Aussage zu produzieren.

2. [EN] *(If no automation tools adopted yet)*
Has the introduction of software automation tools ever been considered in your company so far? If so, which ones?

[DE] *(Wenn bisher noch keine Automatisierungstools in Verwendung)*
Wurde die Einführung von softwarebasierten Automatisierungstools in Ihrem Unternehmen jemals in Betracht gezogen? Wenn ja, welche?

2.1 [EN] **Can you think of workflows in a newsroom that would be well suited to be supported or taken over by intelligent technologies?**

[DE] **Fallen Ihnen Workflows in der Redaktion ein, die sich eignen würden, von intelligenten Technologien unterstützt oder ersetzt zu werden?**

3. [EN] *(If no automated news generation adopted yet)*
In your opinion, what are the key reasons that kept [company name] from adopting automated news generation so far?

[DE] *(Wenn bisher noch keine automatisierte Generierung von News in Verwendung)*
Was sind Ihrer Meinung nach die Hauptgründe, die [Unternehmen] bisher davon abgehalten haben, die automatisierte Generierung von News einzuführen?

4. [EN] **In your opinion, how does [company name] currently differentiate itself from competitors?**

[DE] **Wie differenziert sich [Unternehmen] Ihrer Meinung nach aktuell von Konkurrenten?**

Generell habe ich da einen ganz guten Einblick, weil es da eine Art internationale Vereinigung gibt, wo wir uns regelmäßig zu Innovations- und Technologiethemen austauschen, mit stark journalistischem und redaktionellem Fokus. Dadurch habe ich einen recht guten Einblick, wo unser Unternehmen [verglichen mit anderen] steht, generell vom Innovationsgrad und dem Grad der Digitalisierung von Abläufen. Wir sind hier schon sehr gut unterwegs muss ich sagen. Wir sind sicher ein Unternehmen, das in unserem Land einen sehr guten Ruf genießt, aber auch international von denen, die mit uns arbeiten, und das zu Recht. Weil im Gegensatz zu vielleicht anderen Medien haben wir tatsächlich auch ein Team von Mitarbeitern und Mitarbeiterinnen, die der Ansicht sind, dass es unsere Aufgabe ist, innovativ zu sein und auszuprobieren was es im Technologiebereich Neues gibt und was es uns bringen kann. Bringen im Sinne von unseren Job noch besser, schneller und einfacher zu machen.

Konkret im Bereich der Textautomatisierung denke ich sind wir im deutschsprachigen Raum wirklich führend. Andere Nachrichtenorganisationen haben auch Lösungen für konkrete Anwendungsfälle, wie Wahlberichterstattung, haben aber nicht so ganz diese weiterführende strategische Ausrichtung, was man damit noch alles machen kann. Im internationalen Bereich gibt es natürlich große Nachrichtenorganisationen, die damit sogar schon um einiges weiter sind oder das schon länger im Einsatz haben.

109

4.1 [EN] **With an extensive adoption of automated journalism in the future, how might this current key differentiating characteristics be subject to change?**

[DE] **Mit einer zukünftig umfassenden Einführung von Automated Journalism, wie könnten sich diese aktuellen Hauptdifferenzierungsmerkmale künftig verändern?**

Das ist schwer zu sagen. Die Automatisierung von journalistischen Texten ist eigentlich nur ein relativ schmales Segment eines generellen Trends, der eben auch die Medienbranche erreicht hat, nämlich Abläufe zu automatisieren. In der Medienbranche geht es hier darum die Informationen zu finden, zu suchen und danach die Geschichte zu distribuieren. Das sind alles Stadien in der Medienproduktion, die bis zu einem gewissen Grad schon jetzt automatisiert unterstützt oder teilautomatisiert sind, oder es zukünftig sicher sein werden. Aber so wie andere Unternehmen auch, würde zum Beispiel kein Autobauer entscheiden, ab sofort ihr gesamtes Tun zu automatisieren. Man wird immer noch gewisse Abteilungen brauchen. Man würde auch nicht die gesamte Fertigung automatisieren und man wird immer Menschen brauchen, die diese Automatisierung konfigurieren. Das heißt, bis zu welchem Grad man im Unternehmen welche Prozesse automatisiert, das wird wahrscheinlich notwendige Strategiearbeit für jedes Medium sein – bis hin zu Fragen wie, was macht das mit unseren Qualitätsstandards? Oder, wie passt das in unsere Qualitätsstandards? Weise ich automatisierten Content aus? Sage ich dem User, dass er diesen Content nur deswegen sieht, weil er vor zwei Tagen irgendetwas anderes angeklickt hat? (Personalisierung – auch eine Fassette der Automatisierung). Also das muss jedes Medienhaus für sich sehr individuell entscheiden, weil die auch unterschiedliche Geschäftsmodelle haben und unterschiedliche Ansprüche an sich und an ihr Publikum. Also es ist nicht einfach, nicht simpel zu beantworten.

Ich habe das Gefühl, dass bei regelbasierter Content Produktion, mit NLG und Variablen, bald einmal das Ende der Fahnenstange erreicht ist. Also für mich persönlich, nachdem ich das jetzt schon ein paar Jahre mache, ist da in Wirklichkeit relativ viel Innovationsluft draußen. Das Ding ist, das kann was es kann, da werden die Sprachmodelle dahinter wohl noch immer besser werden, aber das Verfahren per se, das steht. Das ist ja auch nichts wirklich Neues und es ist zwar ein bisschen komplexer, aber jetzt nicht „Rocket Science".

Ich glaube der nächste Schritt wird es tatsächlich sein, was mit „echtem" Machine Learning machbar ist, im Journalismus und im Bereich Content. Da gibt es ja auch schon einige Ansätze, überwiegend im Bereich Prototyping und Proof-of-Concepts. Ein Beispiel ist hier das JournalismAI Festival, wo man schon ein paar Showcases von einzelnen Medien sehen kann. Hier sind wir zum Glück jetzt gerade so weit, dass der erste Hype, im Sinne von „Wow, diese Maschine schreibt einen Artikel wie ein Mensch und man braucht den Menschen

110

nicht mehr", abgeflacht ist. Jetzt beginnt die Arbeit, wo ich diese Technologie verwenden kann, um mir meinen Redaktionsassistenten zu trainieren, der mir ganz spezifische typische „tedious Tasks", also die wirklich langweiligen Arbeiten, abnimmt oder mich dabei unterstützt. Und das ist eigentlich der nächste Schritt. Da bin ich schon sehr gespannt, was sich tun wird. Das ist auch etwas, das wir als nächstes angehen werden und auch angehen müssen, finde ich.

Interview G

1. [EN] **Does [company name] currently use intelligent technologies for the automation of workflows, particularly automated news generation? If so, which ones?**

   [DE] **Nutzt [Unternehmen] aktuell intelligente Technologien für die Automation von Workflows, insbesondere zur automatischen Generierung von News? Wenn ja, welche?**

   Es gibt zwei bis drei Projekte. Das eine ist die Generierung von Fußballberichten in den tiefen Ligen. Das zweite Projekt ist die Abstimmungsberichterstattung. Also wenn bei uns Abstimmungen auf nationaler Ebene stattfinden, dann generieren wir Texte für jede Gemeinde, wie dort abgestimmt wurde. Das sind also unsere ständigen Systeme, die auch vollautomatisch laufen und nicht mehr groß irgendwelche manuellen Zwischenschritte benötigen. Hier findet auch ein ähnliches Texttemplate Verwendung. Der dritte Bereich ist, dass immer wenn wir Daten haben, welche es für alle Gemeinden gibt, dann generieren wir damit eine Story. Dann schreiben wir ein Texttemplate, mit dem die Texte für die Gemeinden generiert werden. Das können beispielsweise Daten zur Altersstruktur sein, womit dann Vergleiche zwischen Gemeinden möglich sind.

   1.1 [EN] *(1.1-1.4: If technology for automation is in use)*
   Is the software that is in use developed by your company itself or is it licensed from a third-party supplier?

   [DE] *(1.1-1.4: Wenn Technologie zur Automatisierung in Verwendung)*
   Ist die verwendete Software eine Eigenentwicklung Ihres Unternehmens oder wird sie von einem Drittanbieter lizensiert?

   Das unterscheidet sich. Das sind immer unterschiedliche Komponenten. Ich kann ein Beispiel geben: Bei den Abstimmungen, da ist mittlerweile eigentlich alles selbst programmiert, von der Datenbeschaffung, über die Datenverarbeitung, bis hin zur Textgenerierung und Auslieferung der Texte an das Frontend. Das ist alles Python, das haben wir selbst programmiert. Also da verwenden wir keine externe Software in der Regel.
   Bei der Fußballberichterstattung, da arbeiten wir mit Retresco zusammen. Aber auch hier gibt es unterschiedliche Komponenten. Die Datenbeschaffung von den Fußballverbänden machen wir selbst. Dann schicken wir die Daten an Retresco und die Textgenerierungskomponente ist dann bei Retresco. Die Speicherung der Daten, die Auslieferung an das Frontend, etc., etc., und auch das Frontend selbst, also wo man sich die Texte dann anschauen kann, das ist wiederum alles von uns programmiert. Diese Mischlösung macht auch Sinn, weil gewisse Sachen kann man selbst besser und andere Sachen können andere besser. Beim Fußball ist es sicher so, dass Retresco eine große Expertise hat. Wir sind mit den Texten sehr zufrieden.

112

Bei unseren ad-hoc Projekten, da machen wir die Texttemplates mit „Arria NLG".

1.1.1 [EN] *(If licensed)*
Was it an option to start developing one's own software? Why (not)?
[DE] *(Wenn lizensiert)*
War es eine Option, die Software selbst zu entwickeln? Warum (nicht)?

1.1.2 [EN] *(If self-developed)*
What were the reasons for an in-house development? What benefits and challenges emerged from this?
[DE] *(Wenn Eigenentwicklung)*
Was waren die Gründe für eine Eigenentwicklung? Welche Vorteile und Herausforderungen haben sich dadurch ergeben?

Die Vorteile sind sicher, dass man selbst alles unter Kontrolle hat, dass man es anpassen kann, wenn man will, dass man, wenn es Fehler gibt, sofort reagieren kann und nicht wochenlang warten muss - der Kundensupport ist je nach Anbieter auch nicht immer der beste.

Die Herausforderung ist, dass eine Eigenentwicklung natürlich immer aufwendig ist.

1.2 [EN] **From your current perspective, what is your company's biggest challenge when it comes to making one's own software solutions?**

[DE] **Was ist aus heutiger Sicht die größte Herausforderung für Ihr Unternehmen, wenn es um die Entwicklung eigener Softwarelösungen geht?**

Ganz generell ist IT eine komplexe Sache und sehr aufwendig. Da ist viel Aufwand dahinter und die Ressourcen sind immer knapp und wir sind nicht Google.

1.3 [EN] **What are the biggest benefits that come along with the software solution deployed in your company?**

[DE] **Was sind die größten Vorteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?**

Wir können den Leserinnen und Lesern einen Service anbieten, den sie so auf anderen Medienportalen nicht finden. Wobei das natürlich auch andere Medien im Land machen. Allerdings ist der Markt relativ gut aufgeteilt, sodass die Konkurrenz in unseren Gebieten nicht so riesig ist. Konkurrierende Medien haben ihren Fokus primär auf anderen Gemeinden, sodass es hier eigentlich fast keine Überschneidung gibt, weswegen wir da auch ein gutes Angebot haben. Es skaliert auch. Also wenn das System einmal funktioniert, dann wird dieser Content generiert, ohne dass man wirklich etwas machen muss.

1.4 [EN] **What are the biggest challenges or downsides that come along with the software solution deployed in your company?**

[DE] Was sind die größten Herausforderungen oder Nachteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?

Die Herausforderung ist, dass diese Systeme einfach funktionieren müssen und robust sein müssen. Zum Beispiel wenn man Daten scrapt, wie wir es beim Fußball machen, dann kann immer etwas schiefgehen. Beispielsweise wenn die Fußballverbände ihre Website abändern, dann hat unser Scraper Probleme und stürzt ab. Man muss wirklich aufpassen, dass die Fehler auch geloggt werden und man auch benachrichtigt wird, wenn irgendetwas schiefläuft. Man muss hier mit Vorsicht herangehen und die Systeme richtig resilient bauen, sonst hat man halt ständig Ausfälle, sodass sich dann wahrscheinlich die Akzeptanz bei Lesern verringern würde.

2. [EN] *(If no automation tools adopted yet)*
   Has the introduction of software automation tools ever been considered in your company so far? If so, which ones?

   [DE] *(Wenn bisher noch keine Automatisierungstools in Verwendung)*
   Wurde die Einführung von softwarebasierten Automatisierungstools in Ihrem Unternehmen jemals in Betracht gezogen? Wenn ja, welche?

   2.1 [EN] **Can you think of workflows in a newsroom that would be well suited to be supported or taken over by intelligent technologies?**

   [DE] **Fallen Ihnen Workflows in der Redaktion ein, die sich eignen würden, von intelligenten Technologien unterstützt oder ersetzt zu werden?**

3. [EN] *(If no automated news generation adopted yet)*
   In your opinion, what are the key reasons that kept [company name] from adopting automated news generation so far?

   [DE] *(Wenn bisher noch keine automatisierte Generierung von News in Verwendung)*
   Was sind Ihrer Meinung nach die Hauptgründe, die [Unternehmen] bisher davon abgehalten haben, die automatisierte Generierung von News einzuführen?

4. [EN] **In your opinion, how does [company name] currently differentiate itself from competitors?**

   [DE] **Wie differenziert sich [Unternehmen] Ihrer Meinung nach aktuell von Konkurrenten?**

Bei uns sind eher die großen Player in solchen Automatisierungsgeschichten involviert. Andere Medienhäuser sind teilweise ein bisschen stärker im Lokaljournalismus verankert. Aber wir haben wie gesagt andere Gebiete, was sich historisch so ergeben hat. Ich denke wir sind unserer Konkurrenz ebenbürtig. Es gibt hier keine großen Unterschiede in der Qualität. Der Unterschied zu den Kleinen ist, dass die

114

Kleinen sich das meist nicht leisten. Weil wie gesagt, das sind IT-Projekte. Das kann man nicht einfach so an einem Nachmittag programmieren.

Abseits von Automated Journalism investieren wir generell recht viel in Technologie. Für uns ist es wichtig, dass unsere Technologien und Plattformen state-of-the-art sind für unsere Nutzerinnen und Nutzer. Aber auch die Systeme, die wir intern für Publikationsprozesse verwenden, halten wir state-of-the-art. Also da wird recht viel investiert. Wir sehen uns eigentlich auch als Technologieunternehmen. Aber ehrlich gesagt, das ist bei anderen wahrscheinlich auch nicht anders, also dass die sich mittlerweile auch als Technologieunternehmen sehen.

4.1 [EN] **With an extensive adoption of automated journalism in the future, how might this current key differentiating characteristics be subject to change?**

[DE] **Mit einer zukünftig umfassenden Einführung von Automated Journalism, wie könnten sich diese aktuellen Hauptdifferenzierungsmerkmale künftig verändern?**

Wahrscheinlich oder hoffentlich führt das dann zu mehr Vielfalt, zu mehr Innovation und zu besseren Ideen, weil der Konkurrenzdruck größer ist. Aber wie das jetzt inhaltlich konkret aussehen könnte, kann ich jetzt auch nicht groß spekulieren. Die Daten werden wahrscheinlich die gleichen sein. Wir verwenden in der Regel öffentlich verfügbare Daten und darauf haben ja alle Zugriff. Kleine Player würden dann wahrscheinlich auf die gleichen Daten zugreifen. Grundsätzlich ist es halt eine sehr langsame Entwicklung. Aber die Entwicklung geht immer mehr dahin, dass mehr Daten öffentlich werden und auch immer mehr Daten erhoben werden - das eröffnet natürlich gleichzeitig immer mehr Möglichkeiten für Automated Journalism. Wenn es mehr Nachfrage nach guten Daten gibt, weil die kleinen Player auch mehr solche Dinge tun, dann profitieren alle davon.

115

Interview H - answered in writing

1. [EN] **Does [company name] currently use intelligent technologies for the automation of workflows, particularly automated news generation? If so, which ones?**

   [DE] **Nutzt [Unternehmen] aktuell intelligente Technologien für die Automation von Workflows, insbesondere zur automatischen Generierung von News? Wenn ja, welche?**

   [Unternehmen] nutzt keine automatischen Tools zur Generierung von News.

   1.1 [EN] *(1.1-1.4: If technology for automation is in use)*
   Is the software that is in use developed by your company itself or is it licensed from a third-party supplier?

   [DE] *(1.1-1.4: Wenn Technologie zur Automatisierung in Verwendung)*
   Ist die verwendete Software eine Eigenentwicklung Ihres Unternehmens oder wird sie von einem Drittanbieter lizensiert?

   1.1.1 [EN] *(If licensed)*
   Was it an option to start developing one's own software? Why (not)?
   [DE] *(Wenn lizensiert)*
   War es eine Option, die Software selbst zu entwickeln? Warum (nicht)?

   1.1.2 [EN] *(If self-developed)*
   What were the reasons for an in-house development? What benefits and challenges emerged from this?
   [DE] *(Wenn Eigenentwicklung)*
   Was waren die Gründe für eine Eigenentwicklung? Welche Vorteile und Herausforderungen haben sich dadurch ergeben?

   1.2 [EN] **From your current perspective, what is your company's biggest challenge when it comes to making one's own software solutions?**

   [DE] **Was ist aus heutiger Sicht die größte Herausforderung für Ihr Unternehmen, wenn es um die Entwicklung eigener Softwarelösungen geht?**

   Entscheidung Make or Buy.

   1.3 [EN] **What are the biggest benefits that come along with the software solution deployed in your company?**

   [DE] **Was sind die größten Vorteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?**

   1.4 [EN] **What are the biggest challenges or downsides that come along with the software solution deployed in your company?**

   [DE] **Was sind die größten Herausforderungen oder Nachteile, die mit der in ihrem Unternehmen verwendeten Software einhergehen?**

2. [EN] *(If no automation tools adopted yet)*
Has the introduction of software automation tools ever been considered in your company so far? If so, which ones?

[DE] *(Wenn bisher noch keine Automatisierungstools in Verwendung)*
Wurde die Einführung von softwarebasierten Automatisierungstools in Ihrem Unternehmen jemals in Betracht gezogen? Wenn ja, welche?

Wir haben sehr viel automatisiert, aber nicht die journalistische Kernarbeit.

> 2.1 [EN] **Can you think of workflows in a newsroom that would be well suited to be supported or taken over by intelligent technologies?**
>
> [DE] **Fallen Ihnen Workflows in der Redaktion ein, die sich eignen würden, von intelligenten Technologien unterstützt oder ersetzt zu werden?**
>
> Wir haben einen Algorithmus entwickelt, der die Homepage bespielt - dies aber nur auf ausgewählten Positionen. Die wichtigsten Positionen bespielt weiterhin das Team der digitalen Blattmacher.

3. [EN] *(If no automated news generation adopted yet)*
In your opinion, what are the key reasons that kept [company name] from adopting automated news generation so far?

[DE] *(Wenn bisher noch keine automatisierte Generierung von News in Verwendung)*
Was sind Ihrer Meinung nach die Hauptgründe, die [Unternehmen] bisher davon abgehalten haben, die automatisierte Generierung von News einzuführen?

Wir haben im Sportbereich darüber nachgedacht, Ticker zu automatisieren. Es braucht aber einen relativ großen Initialaufwand, und davor sind wir bisher zurückgeschreckt. Auch ist es ein Zeitproblem, da wir sehr viele Projekte parallel haben.

4. [EN] **In your opinion, how does [company name] currently differentiate itself from competitors?**

[DE] **Wie differenziert sich [Unternehmen] Ihrer Meinung nach aktuell von Konkurrenten?**

[Unternehmen] ist das führende Portal für News und Unterhaltung in [Land]. Unsere USP sind im Bereich Sport, Meinung, TV. [Unternehmen] ist dabei, wenn etwas passiert - und die Nutzer kommen zu [Unternehmen], wenn etwas passiert.

> 4.1 [EN] **With an extensive adoption of automated journalism in the future, how might this current key differentiating characteristics be subject to change?**
>
> [DE] **Mit einer zukünftig umfassenden Einführung von Automated Journalism, wie könnten sich diese aktuellen Hauptdifferenzierungsmerkmale künftig verändern?**

Ich kann mir nicht vorstellen, dass wir Automated Journalism umfassend einführen werden in der nächsten Zeit.

Bibliography

Adams, William. "Conducting Semi-Structured Interviews". In: *Handbook of Practical Program Evaluation*. 4th ed. 2015. Chap. 19, pp. 492–505.

AIPPI World Congress. *APPI. Resolution on copyright in artificially generated works*. Tech. rep. London: AIPPI World Congress, 2019, pp. 1–4.

Akter, Shahriar et al. "Algorithmic bias in data-driven innovation in the age of AI". In: *International Journal of Information Management* 60 (2021).

Ali, Waleed and Mohamed Hassoun. "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities". In: *International Journal of Media, Journalism and Mass Communications (IJMJMC)* 5.1 (2019), pp. 40–49.

Ananny, Mike and Kate Crawford. "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability". In: *New Media & Society* 20.3 (2018), pp. 973–989.

Anderson, C W. "Notes Towards an Analysis of Computational Journalism". In: *HIIG Discussion Paper Series No. 2012-1*. 2011.

— "Towards a sociology of computational and algorithmic journalism". In: *New Media and Society* 15.7 (2012), pp. 1005–1021.

Association for Computing Machinery US Public Policy Council (USACM). "Statement on algorithmic transparency and accountability". In: *USACM press releases* January 12 (2017), pp. 1–2.

Biswal, Santosh Kumar and Nikhil Kumar Gouda. "Artificial Intelligence in Journalism: A Boon or Bane?" In: *Optimization in Machine Learning and Applications*. Ed. by Anand J Kulkarni and Suresh Chandra Satapathy. Singapore: Springer Nature Singapore Pte Ltd, 2020, pp. 155–167.

Boumans, Jelle W and Damian Trilling. "Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars". In: *Digital Journalism* 4.1 (2016), pp. 8–23.

Broussard, Meredith et al. "Artificial Intelligence and Journalism". In: *Journalism & Mass Communication Quarterly* 96.3 (2019), pp. 673–695.

Bucher, Taina. "'Machines don't have instincts': Articulating the computational in journalism". In: *New Media & Society* 19.6 (2017), pp. 918–933.

Burrell, Jenna. "How the machine 'thinks': Understanding opacity in machine learning algorithms". In: *Big Data & Society* 3.1 (2016).

Carlson, Matt. "Automated journalism: A posthuman future for digital news?" In: *The Routledge Companion to Digital Journalism Studies*. Taylor and Francis, 2016, pp. 226–234.

— "The Robotic Reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority". In: *Digital Journalism* 3.3 (2015), pp. 416–431.

Caswell, David. "Structured Journalism and the Semantic Units of News". In: *Digital Journalism* 7.8 (2019), pp. 1134–1156.

Caswell, David and C W Anderson. "Computational Journalism". In: *The International Encyclopedia of Journalism Studies*. American Cancer Society, 2019, pp. 1–8.

Caswell, David and Konstantin Nicholas Dörr. "Automated Journalism 2.0: Event-driven narratives: From simple descriptions to real stories". In: *Journalism Practice* 12.4 (2018), pp. 477–496.

Clerwall, Christer. "Enter the Robot Journalist: Users' perceptions of automated content". In: *Journalism Practice* 8.5 (2014), pp. 519–531.

Coddington, Mark. "Clarifying Journalism's Quantitative Turn". In: *Digital Journalism* 3.3 (2015), pp. 331–348.

Cramer, Henriette et al. "Assessing and Addressing Algorithmic Bias in Practice". In: *Interactions* 25.6 (Oct. 2018), pp. 58–63.

Cremaschi, Marco et al. "Supporting journalism by combining neural language generation and knowledge graphs". In: *CEUR Workshop Proceedings*. Ed. by Semeraro G Bernardi R. Navigli R. Vol. 2481. CEUR-WS, 2019.

Dalen, Arjen van. "The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists". In: *Journalism Practice* 6.5-6 (2012), pp. 648–658.

Dalgali, Ayse and Kevin Crowston. "Algorithmic Journalism and Its Impacts on Work". In: *Computation + Journalism Symposium (2020)* (2020), pp. 1–5.

Danks, David and Alex John London. "Algorithmic Bias in Autonomous Systems." In: *IJCAI*. Vol. 17. 2017, pp. 4691–4697.

Diakopoulos, Nicholas. "Accountability in Algorithmic Decision-Making: A View from Computational Journalism". In: *Queue* 13.9 (Nov. 2015), pp. 126–149.

— "Algorithmic Accountability: Journalistic investigation of computational power structures". In: *Digital Journalism* 3.3 (2015), pp. 398–415.

Diakopoulos, Nicholas and Michael Koliska. "Algorithmic Transparency in the News Media". In: *Digital Journalism* 5.7 (2017), pp. 809–828.

Díaz-Noci, Javier. "Artificial intelligence systems-aided news and copyright: Assessing legal implications for journalism practices". In: *Future Internet* 12.5 (2020).

Dörr, Konstantin Nicholas. "Mapping the field of Algorithmic Journalism". In: *Digital Journalism* 4.6 (2016), pp. 700–722.

Dörr, Konstantin Nicholas and Katharina Hollnbuchner. "Ethical Challenges of Algorithmic Journalism". In: *Digital Journalism* 5.4 (2017), pp. 404–419.

Elikci, Cigdem. "Integration of Artificial Intelligence Into Austrian Journalism: A Potential Threat to Austrian Journalism?" In: *Analyzing Future Applications of AI, Sensors,*

120

and Robotics in Society. Ed. by Thomas Heinrich Musiolik and Adrian David Cheok. Hershey, PA, USA: IGI Global, 2021, pp. 182–200.

Fanta, Alexander. "Putting Europe's robots on the map: automated journalism in news agencies". In: *Reuters Institute Fellowship Paper* (2017).

Fazelpour, Sina and David Danks. "Algorithmic bias: Senses, sources, solutions". In: *Philosophy Compass* 16.8 (2021).

Flew, Terry et al. "The promise of computational journalism". In: *Journalism Practice* 6.2 (2012), pp. 157–171.

Freialdenhoven, Jan et al. "Do AI generated works qualify for copyright? Summary of the German report to the AIPPI 2019 Study Question on copyright in artificially generated works". In: *IntellectualProperty* 1 (2020), pp. 28–31.

Fu, Runshan, Yan Huang, and Param Vir Singh. "Artificial Intelligence and Algorithmic Bias: Source, Detection, Mitigation, and Implications". In: *INFORMS TutORials in Operations Research.* 2020. Chap. 2, pp. 39–63.

Galily, Yair. "Artificial intelligence and sports journalism: Is it a sweeping change?" In: *Technology in Society* 54 (2018), pp. 47–51.

Gatt, Albert and Emiel Krahmer. "Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation". In: *Journal of Artificial Intelligence Research* 61.1 (Jan. 2018), pp. 65–170.

Graefe, Andreas. *Guide to Automated Journalism.* Tow Center for Digital Journalism, Columbia Journalism School, 2016.

Graefe, Andreas and Nina Bohlken. "Automated journalism: A meta-analysis of readers' perceptions of human-written in comparison to automated news". In: *Media and Communication* 8.3 (2020), pp. 50–59.

Habel, Dominic. *Roboterjournalismus.* 1st ed. Vol. 44. Schriften zum Medien- und Informationsrecht. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG, 2019.

Haim, Mario and Andreas Graefe. "Automated News: Better than expected?" In: *Digital Journalism* 5.8 (2017), pp. 1044–1059.

Hammond, Philip. "From computer-assisted to data-driven: Journalism and Big Data". In: *Journalism* 18.4 (2017), pp. 408–424.

Hansen, Mark et al. *Artificial Intelligence: Practice and Implications for Journalism.* Tech. rep. Columbia Journalism School, 2017, pp. 1–21.

High-Level Expert Group on Artificial Intelligence. *European Ethics Guidelines for Trustworthy Artificial Intelligence (AI).* Tech. rep. European Commission, 2019.

Hunt, Robert and Fenwick McKelvey. "Algorithmic Regulation in Media and Cultural Policy: A Framework to Evaluate Barriers to Accountability". In: *Journal of Information Policy* 9 (Sept. 2019), pp. 307–335.

Kaa, H A J van der and E J Krahmer. "Journalist versus news consumer: The perceived credibility of machine written news". In: *Computation+Journalism conference.* 2014.

Kallio, Hanna et al. "Systematic methodological review: developing a framework for a qualitative semi-structured interview guide". In: *Journal of Advanced Nursing* 72.12 (2016), pp. 2954–2965.

Kaminski, Margot E and Gianclaudio Malgieri. "Multi-Layered Explanations from Algorithmic Impact Assessments in the GDPR". In: *Conference on Fairness, Accountability, and Transparency*. New York, USA: Association for Computing Machinery, 2020, pp. 68–79.

Kaplan, Andreas and Michael Haenlein. "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence". In: *Business Horizons* 62.1 (2019), pp. 15–25.

Kim, Daewon and Seongcheol Kim. "Newspaper companies' determinants in adopting robot journalism". In: *Technological Forecasting and Social Change* 117 (2016), pp. 184–195.

— "Newspaper journalists' attitudes towards robot journalism". In: *Telematics Informatics* 35.2 (2018), pp. 340–357.

Komatsu, Tomoko et al. "AI Should Embody Our Values: Investigating Journalistic Values to Inform AI Technology Design". In: *11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI)*. New York, USA: Association for Computing Machinery, 2020, pp. 1–13.

Kordzadeh, Nima and Maryam Ghasemaghaei. "Algorithmic bias: review, synthesis, and future research directions". In: *European Journal of Information Systems* 31.3 (2022), pp. 388–409.

Kotenidis, Efthimis and Andreas Veglis. "Algorithmic Journalism—Current Applications and Future Perspectives". In: *Journalism and Media* 2.2 (2021), pp. 244–257.

Krawarik, Verena et al. *AI.AT.Media. AI and the Austrian Media Sector: Mapping the Landscape, Setting a Course*. Tech. rep. Vienna: Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie, 2021.

Kunert, Jessica. "Automation in sports reporting: Strategies of data providers, software providers, and media outlets". In: *Media and Communication* 8.3 (2020), pp. 5–15.

— "Journalists, meet your new colleague algorithm: The impact of automation on content distribution and content creation in the newsroom". In: *Media Trust in a Digital World: Communication at Crossroads*. Springer International Publishing, 2019, pp. 135–148.

Laat, Paul B de. "Big Data and Algorithmic Decision-Making: Can Transparency Restore Accountability?" In: *ACM SIGCAS Computers and Society* 47.3 (Sept. 2017), pp. 39–53.

Latar, Noam Lemelshtrich. "Robot journalism". In: *Robot Journalism*. 2018. Chap. 3, pp. 29–40.

— *Robot Journalism: Can Human Journalism Survive?* World Scientific, 2018.

— "The Robot Journalist in the Age of Social Physics: The End of Human Journalism?" In: *The New World of Transitioned Media: Digital Realignment and Industry Transformation*. Ed. by Gali Einav. Cham: Springer International Publishing Switzerland, 2015, pp. 65–80.

Lecompte, Celeste. *Automation in the Newsroom. How algorithms are helping reporters expand coverage, engage audiences, and respond to breaking news. URL: https://niemanreports.org/articles/automation-in-the-newsroom/*. 2015.

Leppänen, Leo, Myriam Munezero, et al. "Data-Driven News Generation for Automated Journalism". In: *10th International Conference on Natural Language Generation (INLG)*. Santiago de Compostela, Spain: Association for Computational Linguistics, 2017, pp. 188–197.

Leppänen, Leo, Hanna Tuulonen, and Stefanie Sirén-Heikel. "Automated journalism as a source of and a diagnostic device for bias in reporting". In: *Media and Communication* 8.3 (2020), pp. 39–49.

Lepri, Bruno et al. "Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges". In: *Philosophy and Technology* 31.4 (2018), pp. 611–627.

Lewis, Seth C, Amy Kristin Sanders, and Casey Carmody. "Libel by Algorithm? Automated Journalism and the Threat of Legal Liability". In: *Journalism & Mass Communication Quarterly* 96.1 (2019), pp. 60–81.

Linden, Carl-Gustav. "Algorithms for journalism: The future of news work". In: *The Journal of Media Innovations* 4 (2017), pp. 60–76.

— "Decades of Automation in the Newsroom: Why are there still so many jobs in journalism?" In: *Digital Journalism* 5.2 (2017), pp. 123–140.

— *News Automation The rewards, risks and realities of 'machine journalism'*. WAN-IFRA, 2019.

— "The bright future of news automation". In: *CEUR Workshop Proceedings*. 2018.

Liu, Bingjie and Lewen Wei. "Machine Authorship In Situ: Effect of news organization and news genre on news credibility". In: *Digital Journalism* 7.5 (2019), pp. 635–657.

Liu, Xiaomo et al. "Reuters Tracer: Toward Automated News Production Using Large Scale Social Media Data". In: *IEEE International Conference on Big Data*. Dec. 2017, pp. 1483–1493.

Luengo, Maria and Susana Herrera-Damas. *News Media Innovation Reconsidered*. John Wiley & Sons, Inc., 2021.

McCracken, Grant. *The Long Interview*. Qualitative Research Methods. SAGE Publications, 1988, p. 88.

Milosavljević, Marko and Igor Vobič. "Human Still in the Loop: Editors Reconsider the Ideals of Professional Journalism Through Automation". In: *Digital Journalism* 7.8 (2019), pp. 1098–1116.

Miroshnichenko, Andrey. "AI to bypass creativity. Will robots replace journalists? (The answer is "yes")". In: *Information* 9.7 (2018).

Montal, Tal and Zvi Reich. "I, Robot. You, Journalist. Who is the Author?: Authorship, bylines and full disclosure in automated journalism". In: *Digital Journalism* 5.7 (2017), pp. 829–849.

— "The death of the author, the rise of the robo-journalist: Authorship, bylines, and full disclosure in automated journalism". In: *The Routledge Handbook of Developments in Digital Journalism Studies*. Taylor and Francis, 2018, pp. 53–63.

Monti, Matteo. "Automated Journalism and Freedom of Information: Ethical and Juridical Problems Related to AI in the Press Field". In: *Opinio Juris in Comparatione* 1 (2018).

Ombelet, Pieter-Jan, Aleksandra Kuczerawy, and Peggy Valcke. "Employing Robot Journalists: Legal Implications, Considerations and Recommendations". In: *25th International Conference Companion on World Wide Web*. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2016, pp. 731–736.

Otterlo, Martijn van. "A machine learning view on profiling". In: *Privacy Due Process and the Computational Turn: The philosophy of law meets the philosophy of technology*. Routledge, 2013.

Saad, Saad and Talat A Issa. "Integration or Replacement: Journalism in the Era of Artificial Intelligence and Robot Journalism". In: *International Journal of Media, Journalism and Mass Communications (IJMJMC)* 6.3 (2020), pp. 1–13.

Salaverría, Ramón. "Digital Journalism". In: *The International Encyclopedia of Journalism Studies*. American Cancer Society, 2019, pp. 1–11.

Schapals, Aljosha Karim and Colin Porlezza. "Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions". In: *Media and Communication* 8.3 (2020), pp. 16–26.

Schell, Katharina. *Whitepaper. Journalistische Textautomatisierung. Status, Potenziale, Limitationen*. Tech. rep. Vienna: APA - Austria Presse Agentur, 2022.

Selbst, Andrew D and Solon Barocas. "The Intuitive Appeal of Explainable Machines". In: *Fordham Law Review* 87.3 (2018), p. 1085.

Shin, Donghee. "Expanding the Role of Trust in the Experience of Algorithmic Journalism: User Sensemaking of Algorithmic Heuristics in Korean Users". In: *Journalism Practice* 16.6 (2022), pp. 1168–1191.

— "Why Does Explainability Matter in News Analytic Systems? Proposing Explainable Analytic Journalism". In: *Journalism Studies* 22.8 (2021), pp. 1047–1065.

Sirén-Heikel, Stefanie et al. "Unboxing news automation: Exploring imagined affordances of automation in news journalism". In: *Nordic Journal of Media Studies* 1.1 (2019), pp. 47–66.

Snyder, Hannah. "Literature review as a research methodology: An overview and guidelines". In: *Journal of Business Research* 104 (2019), pp. 333–339.

Tal, Avital Shulner et al. "'End to End' Towards a Framework for Reducing Biases and Promoting Transparency of Algorithmic Systems". In: *14th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)*. June 2019, pp. 1–6.

Tandoc Jr, Edson C, Lim Jia Yao, and Shangyuan Wu. "Man vs. Machine? The Impact of Algorithm Authorship on News Credibility". In: *Digital Journalism* 8.4 (2020), pp. 548–562.

Tewari, Shubhra et al. "Perceptions of Human and Machine-Generated Articles". In: *Digital Threats: Research and Practice* 2.2 (Apr. 2021).

Thurman, Neil, Konstantin Nicholas Dörr, and Jessica Kunert. "When Reporters Get Hands-on with Robo-Writing: Professionals consider automated journalism's capabilities and consequences". In: *Digital Journalism* 5.10 (2017), pp. 1240–1259.

Torrijos, J L Rojas. "Automated sports coverage. Case study of bot released by The Washington post during the río 2016 and Pyeongchang 2018 olympics". In: *Revista Latina de Comunicacion Social* 74 (2019), pp. 1729–1747.

Utasch, Frederic and Claudia Loebbecke. "Intelligent technologies shaping business models for journalistic content provision: A concept matrix". In: *Americas Conference on Information Systems (AMCIS)*. 2017.

Vázquez-herrero, Jorge et al. *Journalistic Metamorphosis: Media Transformation in the Digital Age*. Vol. 70. Springer Nature Switzerland AG, 2020, pp. 1–213.

Weber, Matthew S and Allie Kosterich. "Coding the News: The role of computer code in filtering and distributing news". In: *Digital Journalism* 6.3 (2018), pp. 310–329.

Weeks, Lin. "Media Law and Copyright Implications of Automated Journalism". In: *NYU Journal of Intellectual Property and Entertainment Law* 4.1 (2014), pp. 67–94.

Wieringa, Maranke. "What to Account for When Accounting for Algorithms: A Systematic Literature Review on Algorithmic Accountability". In: *Conference on Fairness, Accountability, and Transparency*. New York, USA: Association for Computing Machinery, 2020, pp. 1–18.

Wu, Shangyuan, Edson C Tandoc Jr, and Charles T Salmon. "A Field Analysis of Journalism in the Automation Age: Understanding Journalistic Transformations and Struggles Through Structure and Agency". In: *Digital Journalism* 7.4 (2019), pp. 428–446.

— "Journalism Reconfigured: Assessing human–machine relations and the autonomous power of automation in news production". In: *Journalism Studies* 20.10 (2019), pp. 1440–1457.

— "When Journalism and Automation Intersect: Assessing the Influence of the Technological Field on Contemporary Newsrooms". In: *Journalism Practice* 13.10 (2019), pp. 1238–1254.

Wu, Yanfang. "Is Automated Journalistic Writing Less Biased? An Experimental Test of Auto-Written and Human-Written News Stories". In: *Journalism Practice* 14.8 (2020), pp. 1008–1028.

Xu, Zhichao and Xintong Lan. "A Scientometric Review of Automated Journalism: Analysis and Visualization". In: *Journal of Physics: Conference Series* 1684 (Nov. 2020), p. 12127.

Zamith, Rodrigo and Mario Haim. "Algorithmic Actants in Practice, Theory, and Method". In: *Media and Communication* 8.3 (2020), pp. 1–4.

Zydorek, Christoph. "Automatisierte Contentproduktion bei Nachrichten – Automated Journalism". In: *Grundlagen der Medienwirtschaft*. Springer Gabler Wiesbaden, 2018, pp. 87–112.