Check for updates

SYSTEMATIC REVIEW

A systematic review of automated journalism scholarship: guidelines and suggestions for future research

[version 1; peer review: 2 approved]

Samuel Danzon-Chambaud

School of Communications, Dublin City University, Dublin, Ireland

v1 **First published:** 24 Mar 2021, **1**:4
https://doi.org/10.12688/openreseurope.13096.1
Latest published: 24 Mar 2021, **1**:4
https://doi.org/10.12688/openreseurope.13096.1

Abstract

Background: The use of advanced algorithmic techniques is increasingly changing the nature of work for highly trained professionals. In the media industry, one of the technical advancements that often comes under the spotlight is automated journalism, a solution generally understood as the auto generation of journalistic stories through software and algorithms, without any human input except for the initial programming.
Methods: In order to conduct a systematic review of existing empirical research on automated journalism, I analysed a range of variables that can account for the semantical, chronological and geographical features of a selection of academic articles as well as their research methods, theoretical backgrounds and fields of inquiry. I then engaged with and critically assessed the meta-data that I obtained to provide researchers with a good understanding of the main debates dominating the field.
Results: My findings suggest that the expression "automated journalism" should be called into question, that more attention should be devoted to non-English speaking scholarship, that the collective and individual impacts of the technology on media practitioners should be better documented and that well-established sociological theories such as institutionalism and Bourdieu's field theory could constitute two adequate frameworks to study automated journalism practices.
Conclusions: This systematic literature therefore provides researchers with an overview of the main challenges and debates that are occurring within the field of automated journalism studies. Future studies should, in particular, make use of institutionalism and field theory to explore how automated journalism is impacting the work of media practitioners, which could help unearth common patterns across media organisations.

Keywords

automated journalism, robot journalism, algorithmic journalism, computational journalism, natural language generation, automation, journalism

Open Peer Review

Approval Status ✓ ✓

	1	2
version 1 24 Mar 2021	✓ view	✓ view

1. **Colin Porlezza**, Università della Svizzera italiana, Lugano, Switzerland

2. **Jessica Kunert**, University of Hamburg, Hamburg, Germany

Any reports and responses or comments on the article can be found at the end of the article.



This article is included in the Sociology gateway.



This article is included in the Marie-Sklodowska-Curie Actions (MSCA) gateway.



This article is included in the Horizon 2020 gateway.



This article is included in the The Future of Work collection.



This article is included in the Artificial Intelligence and the Social Sciences and Humanities collection.

Corresponding author: Samuel Danzon-Chambaud (samuel.danzonchambaud2@mail.dcu.ie)

Author roles: Danzon-Chambaud S: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No [765140], (JOLT).

Copyright: © 2021 Danzon-Chambaud S. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Danzon-Chambaud S. **A systematic review of automated journalism scholarship: guidelines and suggestions for future research [version 1; peer review: 2 approved]** Open Research Europe 2021, **1**:4 https://doi.org/10.12688/openreseurope.13096.1

First published: 24 Mar 2021, **1**:4 https://doi.org/10.12688/openreseurope.13096.1

Introduction

The use of advanced algorithmic techniques is increasingly changing the nature of work for highly trained professionals. In the medical industry, health professionals can rely on neural networks to detect skin cancers among patients (Esteva *et al.*, 2017) while advanced classification or natural language processing methods can be deployed to predict judgements in the legal industry (Katz *et al.*, 2017; Medvedeva *et al.*, 2020). In the media industry, one of the technical advancements that often comes under the spotlight is automated journalism, a solution generally understood as the auto generation of journalistic stories through software and algorithms, with no human intervention except for the initial programming (Graefe, 2016). Also popularised under the terms "robot journalism" (Lemelshtrich Latar, 2018), "algorithmic journalism" (Dörr, 2016) and "machine-written news" (van Dalen, 2012), automated journalism builds on natural language generation (NLG), a computer process that triggers text generation. Employed for several decades in domains such as sports, finances and weather forecasting (Dörr, 2016), NLG systems traditionally involve pre-written templates that are filled through a set of specific rules, while other methods rely on machine learning techniques to "learn patterns of language use from large corpora of examples" (Diakopoulos, 2019, p. 101; Graefe, 2016).

Automated journalism started to be more widely discussed in the 2010s as The Los Angeles Times started covering homicides in an automated fashion and launched a tool to generate earthquake alerts, while The Associated Press partnered with the firm Automated Insights to automate the majority of its corporate earnings stories. At the same time, The Washington Post developed an in-house software to produce short automated stories and alerts during the 2016 Rio Olympics. In France, Le Monde collaborated with the firm Syllabs to automatically cover the results of the 2015 regional elections while, in Switzerland, Tamedia Group used Automated Insights' solution to report on the outcome of a 2018 referendum (Plattner & Orel, 2019). The BBC also resorted to a form of semi-automated journalism to cover the results of the 2019 United Kingdom general election. In addition to this, a dozen of European news agencies adopted or planned for the development of the technology (Fanta, 2017).

At the same time, automated journalism is also increasingly attracting the attention of the academic community. The algorithmic nature of automated journalism turns it into a prime candidate for the study of algorithmic decision-making, an area of research that looks into tasks and processes commanded by algorithms. As such, the technology is sometimes discussed along with algorithmic distribution of media content (Carlson, 2018; Napoli, 2014). Automated journalism can also be investigated as being part of computational journalism studies (Bucher, 2017; Coddington, 2015), a discipline that initially focused on the use of advanced software to assist journalists with their daily workflow (Cohen *et al.*, 2011; Flew *et al.*, 2012; Hamilton & Turner, 2009) and then expanded to journalists' abilities to solve problems through abstraction and computing skills (Gynnild, 2014; Stavelin, 2013). It is for this reason that automated journalism is sometimes only mentioned as "computational journalism" (Lindén, 2017; Young & Hermida, 2015). Finally, automated journalism can be studied in the context of newsroom automation and artificial intelligence in journalism, a discipline that looks into the latest applications of computational breakthroughs in the media industry, such as platforms handling the distribution of news on social media and data-mining techniques for investigative journalism (Diakopoulos, 2019; Marconi, 2020).

This systematic literature review analyses the key features of a selection of academic articles on automated journalism in order to provide a comprehensive overview of the field, and to contribute guidelines and suggestions for future research endeavours. Although Graefe & Bohlken (2020) conducted a meta-analysis focusing on readers' perceptions of automated journalism, no scholarship has so far assessed the full range of peer-reviewed journal articles contributing empirical evidence to this growing research area. My research questions are therefore as follows:

1. Which patterns arise from a selection of key variables used in a corpus of peer-reviewed articles on automated journalism that are systematically retrieved?

2. Based on this pattern analysis, what are the main debates dominating the field of automated journalism studies?

Methods

Selection of corpus

To carry out this review, I retrieved the documents that constituted the final corpus in a systematic manner, with the following combined search query: "automated journalism" OR "algorithmic journalism" OR "robot journalism" OR "machine-written journalism" OR "computational journalism." I searched through five databases (Taylor and Francis, Sage, ScienceDirect, SpringerLink, and Scopus) as I looked for content published between 2005 and 2020 to represent the last 15 years of research, a range that shall accommodate the purpose of this research as it goes back five years prior to the launch of The Los Angeles Times' pioneering project on automating homicide coverage.

Out of close to 500 results, I only selected articles written in English, that are based on empirical research to assess the latest findings in the field, and that presented an exclusive focus on automated journalism, that I understood as "the process of using software or algorithms to automatically generate news stories without human intervention – after the initial programming of the algorithm" (Graefe, 2016, p. 14), a definition that I also extended to auto generated text so as to account for recent developments in the field (Lindén, 2017). In the end, this final corpus was constituted of 33 scholarly articles on automated journalism (Danzon-Chambaud, 2021a), a number that I am satisfied with, given that automated journalism was introduced into newsrooms and attracted scholars' attention only recently. This number involves almost three times as many articles as in Graefe's and Bohlken's meta-analysis of readers'

perceptions of the technology (2020) and comes close to the 40 articles analysed in an authoritative systematic review of data journalism scholarship (Ausserhofer *et al.*, 2017), which constitutes another major development in the field of media and communication studies.

Then, in order to conduct an efficient review of the field, I looked, first, for variables I could quantitatively measure, such as the semantical, chronological and geographical features of the articles studied as well as their research methods, fields of inquiry and theoretical backgrounds. To do this, I retrieved every keyword mentioned in the corpus, the years the articles were published online as well as the countries they originated from, the domains of investigation and methods used in each article, and lastly, the theoretical considerations and any bibliographic reference cited more than five times in the entire corpus. Second, I engaged with and critically assessed these meta-data in order to equip researchers with a more qualitative understanding of the main debates dominating the field. No risk of bias was found, apart from limiting my search to articles only written in English, a point that I address in the study.

Results

"Robot journalism" vs. "automated journalism"

In order to investigate the various semantics used in the field of automated journalism, I first analysed the different keywords mentioned in the corpus (Figure 1). I found that the most-frequently used ones referred to the terms "robot journalism" and "automated journalism", two expressions that are regularly employed in mainstream media and academia to evoke the computer-generation of news text, but which both face criticisms related to their exact meaning.

The most problematic use of the term "robot journalism" lies in the fact that NLG involves a computer script and no actual

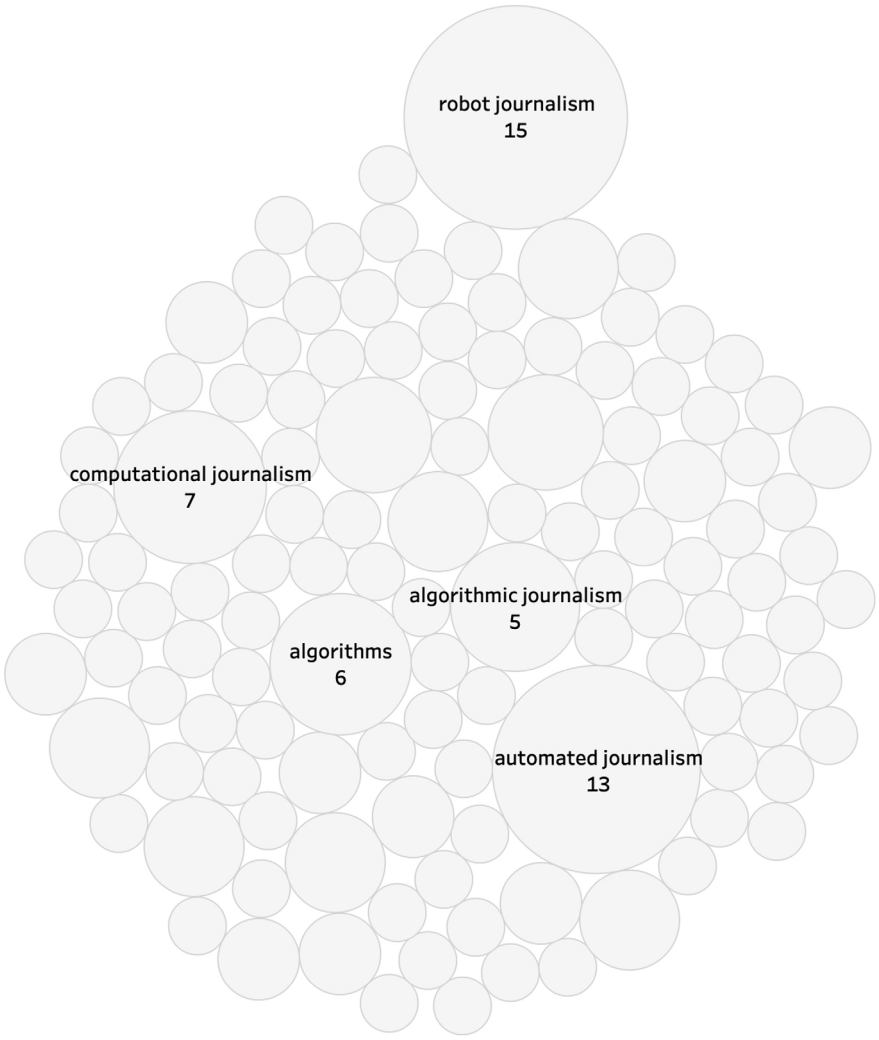


Figure 1. Most-frequent keywords used in corpus. Only keywords mentioned 5 times or more are indicated.

robot (Dörr, 2016). Lindén (2017, p. 125) cautions against a "popular but banal conceptualisation where illustrators more often portray robots writing on computer keyboards," that ultimately plays on journalists' fears of being made redundant and prevents newsrooms from being more innovative (Lindén & Dierickx, 2019). This metaphor could also be detrimental to the acceptance of automated stories, since readers can ultimately feel deceived after being drawn to believe that the technology exhibits a form of humanness (Waddell, 2018). In addition, there could be no added benefit to using this expression since readers have been revealed to be equally receptive to automated news labelled as a product of algorithms or identified through software name (Waddell, 2019a). Only readers that were previously exposed to robots in popular culture seem to perceive this metaphor in a more positive light (Waddell, 2018).

That being said, the use of the alternate expression "automated journalism" could also be up for debate. Although it is increasingly a much-preferred term within industry and academia, it is argued that its focus on computer-generated text is too narrow and does not reflect media practitioners' views. In fact, Wu *et al.* (2019b, p. 1453) advance that the definition of "automated journalism" could be extended to a whole other range of tasks, which could include, for instance, "anything from the machine aggregating and funnelling of content, to data scraping and auto-publication of stories".

If the use of the "robot journalism" metaphor has largely been called into question, the use of the expression "automated journalism" – that I employ here because it is less controversial among scholars – has rarely been interrogated in the same manner. It should nevertheless be reflected upon, since this term could also encompass other algorithmic tasks in journalism such as retrieving newsworthy data in investigative reporting (Broussard, 2015; Stray, 2019) or using algorithms to automate fact-checking (Graves, 2018). Besides, the use of other popular keywords such as "computational journalism" and "algorithmic journalism" could not really supplant "automated journalism" as they do have an equally broad meaning, leaving room for discussion among scholars.

Looking beyond the Anglosphere

The second type of variables I looked for were the years the articles were published online and the countries from where research originated (Table 1). While no publication was found prior to 2012, a steady growth in the number of articles

Table 1. Articles' online publication years and countries of origin.

Country	2012	2013	2014	2015	2016	2017	2018	2019	2020	Grand Total
Australia	–	–	–	–	–	–	–	1	–	**1**
Canada	–	–	1	–	–	–	–	–	–	**1**
Denmark	1	–	–	–	–	–	–	–	–	**1**
Germany	–	–	–	–	1	1	–	–	–	**2**
Finland	–	–	–	–	1	–	1	–	–	**2**
Great Britain	–	–	–	–	–	–	–	1	–	**1**
India	–	–	–	–	–	–	1	–	–	**1**
Israel	–	–	–	–	1	–	–	–	–	**1**
Netherlands	–	–	–	–	–	–	1	–	–	**1**
Singapore	–	–	–	–	–	–	–	–	1	**1**
South Korea	–	–	–	–	1	2	1	–	–	**4**
Spain	–	–	–	–	–	–	–	2	1	**3**
Sweden	–	–	1	–	–	–	–	–	–	**1**
Switzerland	–	–	–	1	–	–	–	–	–	**1**
United States	–	–	1	1	–	2	2	3	–	**9**
China-United States	–	–	–	–	–	–	1	–	–	**1**
Germany-Switzerland	–	–	–	–	–	1	–	–	–	**1**
Switzerland-United States	–	–	–	–	–	1	–	–	–	**1**
Grand Total	**1**	**0**	**3**	**2**	**4**	**7**	**7**	**7**	**2**	**33**

published on automated journalism started to be noticeable from 2014 onwards, with the only exception of a small decrease in 2015 and an incomplete picture in 2020 since I finalised my data collection in mid-June of that year (it is also likely that research published in 2020 will be impacted by the coronavirus outbreak). This growth can be interpreted as a reflection of the adoption of the technology, since major media outlets announced their passage to automated journalism at about the same time. Looking at the countries from which the research originated, the results were not surprising as I limited my search to English-written scholarship; English-speaking countries (i.e. United States, Australia, Canada, United Kingdom) constituted the largest group in my ranking, followed by other western countries (i.e., Germany, Denmark, Finland, Israel, Netherlands, Spain, Sweden, Switzerland) and Asian countries (South Korea, India, Singapore). No scholarship associated with an African country or a South American country were found, but this probably was to do with my focus on scholarship written in English than with technology penetration in developing countries, though this is not negligible. As a matter of fact, the 2019 conference of the International Association for Media and Communication Research featured a few papers on automated journalism from African (Ojomo & Ikem, 2019; Salawu, 2019) and South American (Limia, 2019) countries, which should result, sooner rather than later, in more publications from these areas.

Having made this observation, it is once I combined the online publication dates with the countries of origin that I started to notice some more patterns. First and foremost, two out of four articles published between 2012 and 2014 originated from Northern European countries (i.e., Denmark and Sweden) while the other half came from Canada and the United States. These two articles from Northern Europe could be considered to be pioneering work on automated journalism since they first tackled its impacts on media practitioners (van Dalen, 2012) and the perceptions triggered among audiences (Clerwall, 2014). These findings also suggest that Northern European outlets had a key role to play in the development of automated journalism, although much of the spotlight was on – and remains with – large media organisations in the United States, such as The Los Angeles Times, The Associated Press and The Washington Post. We can look for instance at Danish financial news agency Ritzau and Swedish media group MittMedia, which have both been engaging with automated journalism as early as 2015. Since they either partially or entirely owned the solution they launched (Falk Eriksen, 2018; Lindén & Tuulonen, 2019), Ritzau and MittMedia differ from other media organisations such as Le Monde and The Associated Press, which adopted automated journalism at about the same time but outsourced its development to external NLG providers.

The combination of online publication years and countries of origin also showed a surge of research coming from East, South and Southeast Asia from 2016 onwards. While most of them were concerned with South Korea (Jung *et al.*, 2017; Kim & Kim, 2017; Kim & Kim, 2018; Kim & Lee, 2018), others originated from India (Visvam Devadoss, 2019), Singapore (Tandoc *et al.*, 2020) and partly from China (Zheng *et al.*, 2018). We can assume that this picture only represents the "tip of the iceberg," as it is likely that additional research has been published in local languages I could not read. In the case of China, for instance, only a handful of information is readily available in English; it is known that news agencies Xinhua and Toutiao as well as media group Caixin are resorting to automated journalism, that the technology is used for the same kind of reporting than in Western outlets (i.e., sports, financial news, weather), and that at least one Chinese firm, Tencent, acts as an NLG provider (Dörr, 2016; Lindén & Tuulonen, 2019). This scarcity of materials written in English prevents me from knowing more about the strategies these organisations develop. Little is known about the "Media Brain" project that China's state agency Xinhua launched in 2018. Initially described as "a first-of-its-kind platform in China that brings cloud computing, the Internet of Things, Big Data and AI technology into news production," it has also reinforced suspicions about the way AI could be used to further disseminate propaganda coming from the Chinese Communist Party.

In sum, pioneering work as seen in Northern Europe as well as recent research published in East, South and Southeast Asia should encourage us to look beyond scholarship written in English. A review of academic articles published in local languages as well as possibly collecting extra empirical data in non-English settings would be critical to fully assessing the reach of automated journalism. Although my corpus is also inclusive of two articles in Spanish that were adapted into English (Rojas Torrijos, 2019; Túñez-Lopez *et al.*, 2019), an English version of other articles written in this language is unavailable on the publisher's website (Sánchez Gonzáles & Sánchez González, 2017; Túñez-López *et al.*, 2018). The same goes for articles written in French (Dierickx, 2018; Dierickx, 2019) and Portuguese (Carneiro dos Santos, 2016), all of which would be especially worth examining as they open a research window not only in Europe, but also in the Americas and in Africa.

Reach and practice

In a third step, I looked at the fields of inquiry and at the methods used in the scholarship under study, in order to delineate the research orientations that characterise them (Figure 2). For this purpose, I discerned two main fields of inquiry: first, the reach of automated journalism, which includes studies on the perceptions of news readers (i.e., whether they rank it similarly to human-written content) and those on the wider repercussions of automated journalism, such as its impacts on the legal and financial spheres; second, the practice of automated journalism, which encompasses technically oriented studies looking into its functioning, studies focused on its deployment within news organisations and those focused on implications for media labour and other associated actors[1].

[1] Two articles adopted a mixed approach: Jung *et al.* (2017) evaluated the perceptions of automated news from a readers and from a journalists perspective while Ford & Hutchinson (2019) focused on the public's and media practitioners' reactions to the utilisation of a chatbot.

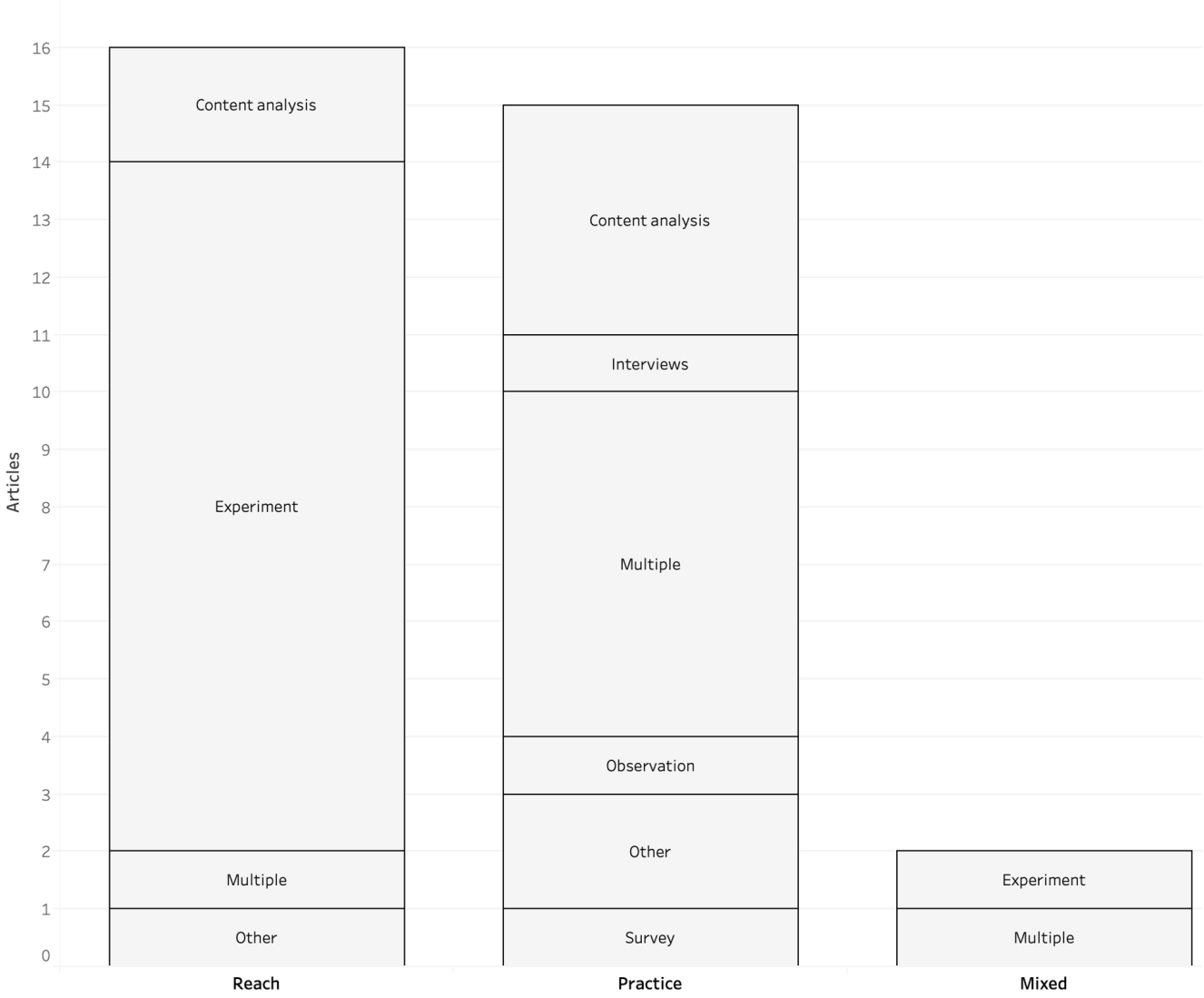


Figure 2. Methods employed for each field of inquiry in corpus.

Regarding the reach of automated journalism, I found that the articles focusing on the perceptions of the technology were almost entirely constituted of experiments conducted on readers. These experiments are either used solely (Clerwall, 2014; Graefe *et al.*, 2018; Haim & Graefe, 2017; Liu & Wei, 2019; Melin *et al.*, 2018; Tandoc *et al.*, 2020; Waddell, 2018; Waddell, 2019a; Waddell, 2019b; Wölker & Powell, 2018; Wu, 2020; Zheng *et al.*, 2018) or in combination with other methods (Kim & Lee, 2018). In line with Graefe's & Bohlken's findings (2020), they highlight in great part that readers evaluate the objectivity, trustworthiness and credibility of automated journalism as similar to those of human journalists, although when reading for pleasure, readers tend to prefer human-written content (Graefe *et al.*, 2018; Haim & Graefe, 2017; Melin *et al.*, 2018; Zheng *et al.*, 2018). Besides, a growing stream of research is looking at combined authorship (humans and algorithms), which has demonstrated promising results so far regarding this method of co-creation (Tandoc *et al.*, 2020; Waddell, 2019a; Waddell, 2019b; Wölker & Powell, 2018).

Besides analysing readers' perceptions, articles focusing on the reach of automated journalism are also concerned with the larger repercussions of the technology. Lewis *et al.* (2019) as well as Díaz-Noci (2020) both resorted to content analysis to analyse the legal impacts of automated journalism. Lewis, Sanders and Carmody argue that media organisations could potentially be condemned for negligence when "defamatory content slips through the cracks" (p. 15). For his part, Díaz-Noci notes that human intervention in the creation of automated

news could help news organisations secure copyright over this type of content. In the business sphere, Blankespoor *et al.* (2018) conducted a series of quantitative tests to analyse the market effects of automated journalism, which showed a correlation between the automation of financial news and an increase in trading volumes for firms less covered prior to automation. However, no impact on determining trade values was found.

In contrast to scholarship concerned with the reach of automated journalism, those focusing on its practice involve many methods, most of the time combined together. I found that technically oriented studies look at the functioning of the technology to demonstrate its potential as well as its limitations. Caswell & Dörr (2018) carried out a series of tests with an NLG software in an attempt to create automated news that would be more complex than existing articles. They hold that reporting on events in a database format (i.e., "structured journalism") makes it possible for uncomplicated stories such as car chases to be automated, but not as far as complicated stories such as parliamentary proceedings are concerned. In another technically oriented article, Visvam Devadoss *et al.*, (2019) managed to create a fully operational NLG system able to draw on materials collected online and on social media to feed a news website.

Another area of investigation pertaining to the practice of automated journalism has to do with organisational impacts on newsrooms. In the realm of sports journalism, two separate content analyses found, on the one hand, that commonalities between automated and human-written output prevailed over differences (Túñez-Lopez *et al.*, 2019) and, on the other hand, that human intervention in the editorial process remained important (Rojas Torrijos, 2019). Regarding the use of "news bots" and "chatbots,"[2] a digital ethnography established that news bots could help media outlets reach out to a niche and geo-specific audience, but sometimes lack data transparency (Lokot & Diakopoulos (2016). Similarly, a content analysis combined with interviews revealed that chatbots can be used to appeal to new audiences, but that they also need to be scrutinised and to be made accountable in order to maintain public media values (Jones & Jones, 2019). In addition, another content analysis combined with interviews showed discrepancies in attribution bylines, which led the authors to suggest a comprehensive attributing policy that covers, on the one hand, stories generated through algorithms only, and on the other hand, news articles co-developed by humans and algorithms (Montal & Reich, 2017).

A third and last domain that relates to automated journalism practices is research that looks at individual impacts on media labour and other associated actors. These studies include first-hand accounts gathered through interviews (Lindén, 2017), surveys (Kim & Kim, 2017) or by combining multiple methods that all include interviews (Dörr, 2016; Kim & Kim, 2018; Thurman *et al.*, 2017; Young & Hermida, 2015). They either highlight a change in newsroom dynamics with established actors such as crime journalists progressively losing their influence to a "new class of computational journalist and non-human journalist" (Young & Hermida, 2015, p. 393), the assumption that journalists will be able to mitigate disruptional changes brought on by automation by showing commitment to their jobs (Lindén, 2017), the conviction among media practitioners that automated journalism is undermined by "fundamental limitations" (Thurman *et al.*, 2017) or the psychological traits journalists adopt when faced with the technology (Kim & Kim, 2018). This stream of research also examines the dominating mindset among media executives, who would reportedly lean toward the implementation of automated journalism instead of hiring additional journalists (Kim & Kim, 2017), and delineates the market position of automated journalism through interviews with NLG providers (Dörr, 2016).

In parallel to these first-hand accounts, two studies engage with qualitative content analysis to examine the collective impacts of automated journalism on media practitioners and other associated actors. In the earliest study on automated journalism found, van Dalen (2012) suggests – after having scrutinised 68 blog posts and newspaper articles – that journalists could react to the introduction of the technology by emphasizing their very human skills (e.g., creativity, personality). He also stresses that routine tasks could be assigned to automated news, so that humans could focus on more demanding formats (provided that newsrooms do not fail to reassign them). Likewise – after having examined 63 pieces of media content, websites and blogs – Carlson (2015) concludes that automated journalism could as much be used to alleviate or augment the work of media practitioners as it could be used to make them redundant. However, these studies investigating broader impacts on media labour and other associated actors were published at an early stage of newsroom adoption of automated journalism, and as such can be considered as being exploratory.

In conclusion, research on the reach of automated journalism showed that readers perceive it similarly to human output for credibility, objectivity and trustworthiness, but not as far as reading for pleasure is concerned. Future research should therefore evaluate the latest breakthroughs in NLG production against this criterion. As for the larger repercussions of automated journalism, it would be worth continuing to look at the influence the technology has on other domains, such as on the legal industry (Monti, 2019; Weeks, 2014). For research dealing with practice, scholarship on the functioning of automated journalism, as well as those focused on organisational impacts, revealed that the technology provides great affordances but should also be weighed against transparency and accountability requirements. That being said, I was unable to draw any definite conclusion with regard to research on individual impacts on media practitioners and other associated actors, while research on the collective impacts of the technology can be considered to be too exploratory. This calls for further research

[2] Although they both belong to the domain of automated journalism, a "news bot" is generally employed to disseminate news on social media in a one-way form of communication while, on the contrary, a "chatbot" is used to generate a conversation between news organisations and their audiences.

into this area, so as to detect common patterns across media organisations.

Underexploited sociological frameworks

Finally, I examined my corpus to see if theory was used, and then compared these results with the fields of inquiry (Figure 3). In most cases, no theory was employed, which is especially valid for studies that focused on the practice of automated journalism, with only three articles actively making use of a theoretical framework (Dörr, 2016; Kim & Kim, 2017; Kim & Kim, 2018) and two building on theory, but not operationalising it in the variables they used (Carlson, 2015; Lindén, 2017). By contrast, two thirds of the articles that made an active use of theory focused on the reach of automated journalism and dealt with evaluating readers' perceptions (Haim & Graefe, 2017; Liu & Wei, 2019; Tandoc *et al.*, 2020; Waddell, 2018; Waddell, 2019a; Waddell, 2019b; Wu, 2020; Zheng *et al.*, 2018).

When comparing the theoretical backgrounds of articles actively using theory with the fields of inquiry (Figure 4), I found that those concerned with readers' perceptions resorted to theories belonging to the realm of psychology (i.e., expectancy violations theory in Liu & Wei, 2019; Waddell, 2018; Waddell, 2019a; Tandoc *et al.*, 2020; MAIN model in Waddell, 2018; Waddell, 2019b; expectation-confirmation theory in Haim & Graefe, 2017; cognitive authority theory in Wu, 2020; similarity attraction in Waddell, 2019b), except for one article using a mixed sociological-psychological framework (i.e., high-context/low-context cultures and holistic/analytic thinking framework in Zheng *et al.*, 2018), while two of the three studies that actively employed theory to analyse the practice of automated journalism used sociological frameworks to investigate it with a focus on media labour (i.e., institutionalism in Dörr, 2016; institutional entrepreneurship, structural inertia, and institutional isomorphism in Kim & Kim, 2017), with the third one using a mixed sociological-psychological

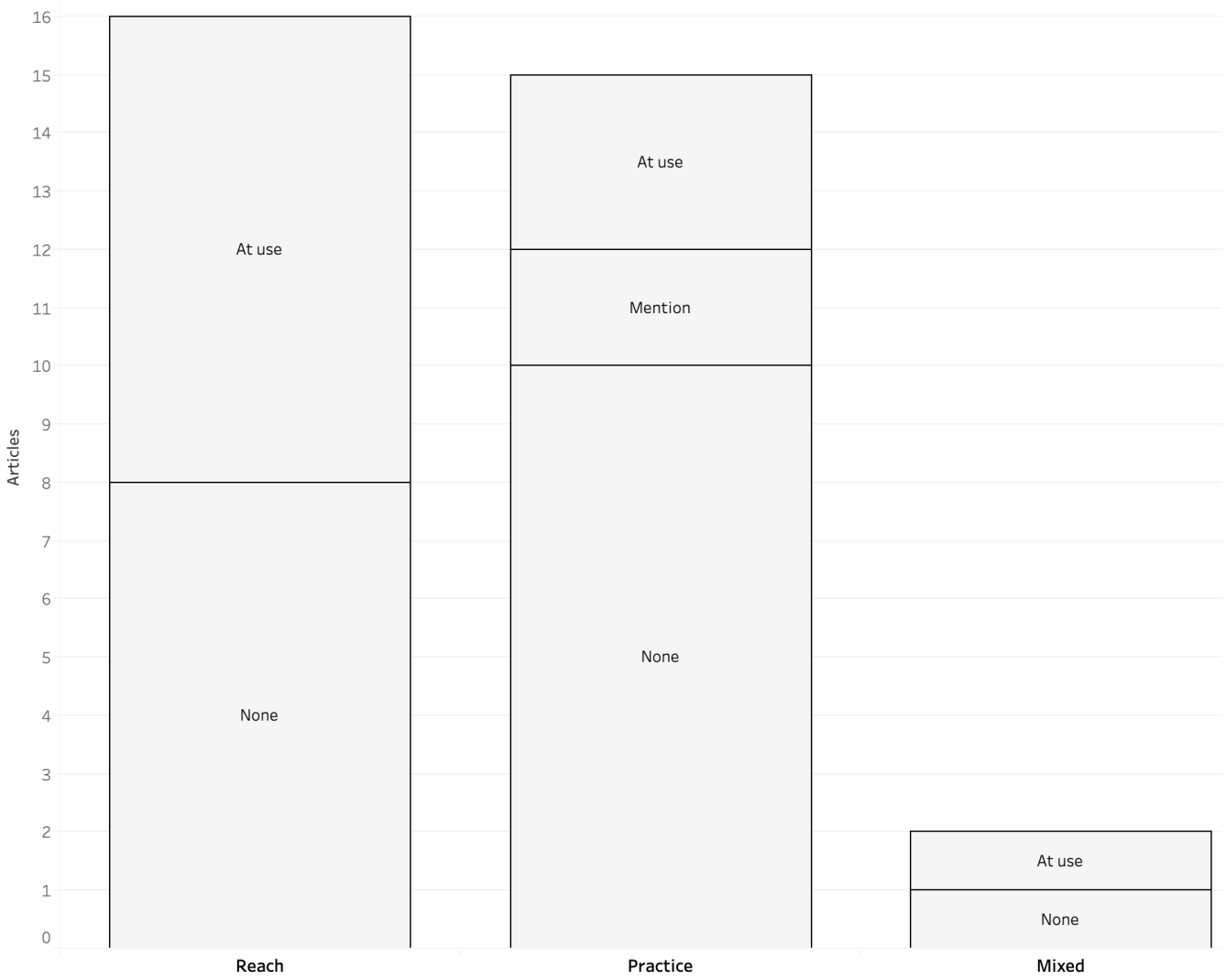


Figure 3. Use of theory for each field of inquiry in corpus.

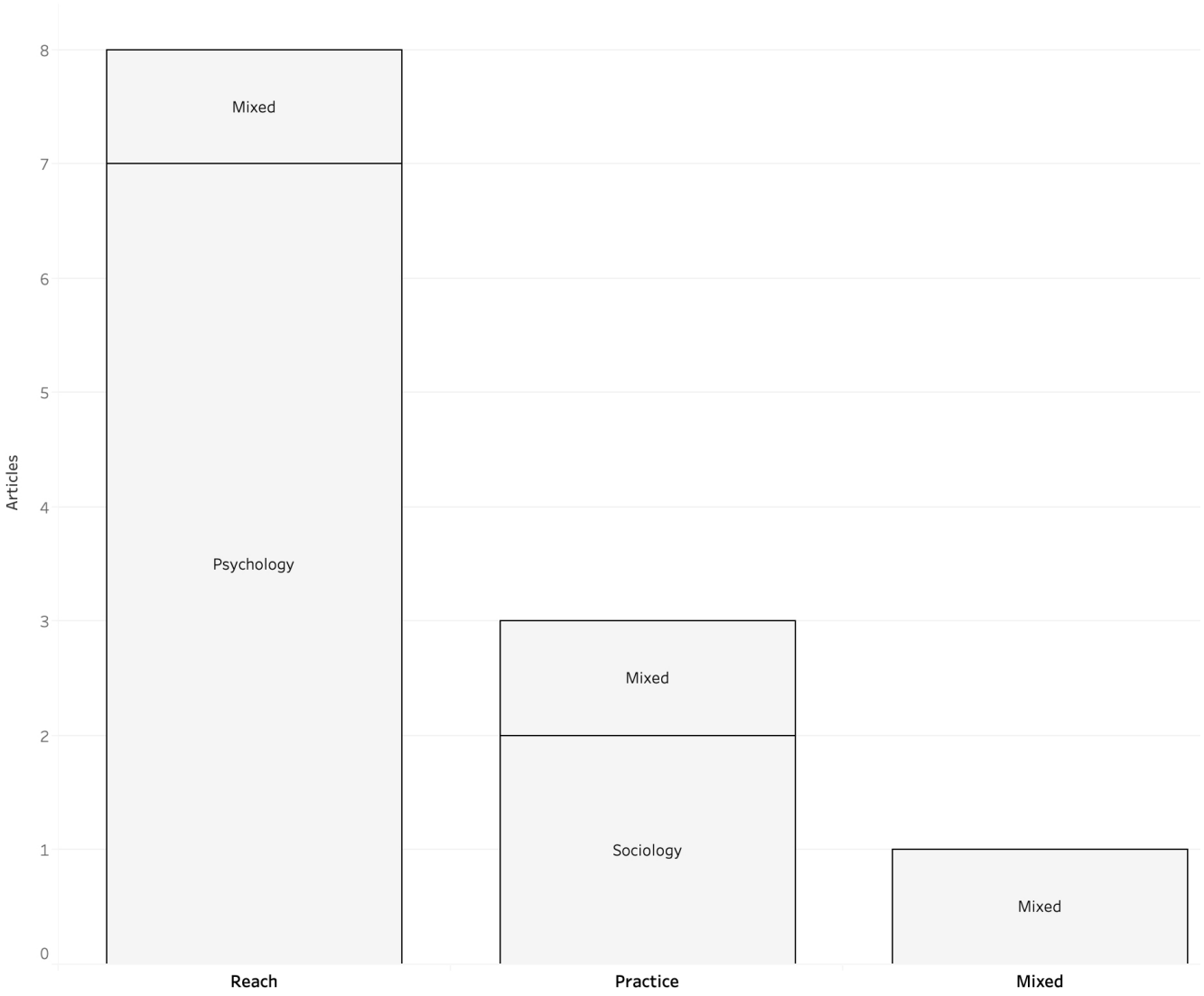


Figure 4. Theoretical backgrounds for each field of inquiry in corpus.

framework for the same purpose (i.e., innovation resistance theory and institutionalism in Kim & Kim, 2018)[3].

In order to have a better idea of the sources that inspired these theoretical considerations, I then looked at the most-cited references throughout the entire corpus (Table 2). First, I noticed that the most-cited sources had to do with empirical studies published on automated journalism, whether they were the same journal articles that I systematically retrieved (Caswell & Dörr, 2018; Carlson, 2015; Clerwall, 2014; Dörr, 2016; Graefe *et al.*, 2018; Jung *et al.*, 2017; Lindén, 2017; Lokot & Diakopoulos, 2016; Montal & Reich, 2017; Thurman *et al.*, 2017; van Dalen, 2012; Waddell, 2018; Young & Hermida, 2015) or other forms of publication that I did not select such as a report (Graefe, 2016), a conference paper (van der Kaa & Krahmer, 2014) and a book chapter (Lemelshtrich Latar, 2015). In addition to this, other empirical materials that did not directly deal with automated journalism, but rather with algorithmic accountability (Diakopoulos, 2015), the use of artificial intelligence for investigative reporting (Broussard, 2015), computational journalism (Karlsen & Stavelin, 2014; Stavelin, 2013) and archival research looking into the evolution of technologically specific forms of work (Powers, 2012) also figured in this listing.

[3] One of the studies focusing both on readers' and journalists' perceptions of automated journalism also used a framework situated at the crossroad of sociology and psychology (In-group and out-group theory in Jung *et al.*, 2017).

Table 2. Most-cited references† in corpus.

Number of citations	Publications
24	Clerwall (2014)
21	Carlson (2015)
20	van Dalen (2012)
17	Anderson (2012)
16	Graefe (2016)
15	Dörr (2016); Young & Hermida (2015)
14	Graefe *et al.* (2018); van der Kaa & Krahmer (2014)
12	Flew *et al.* (2012)
11	Coddington (2015); Diakopoulos (2015); Lemelshtrich Latar (2015)
10	Thurman *et al.* (2017)
9	Gillespie (2014); Napoli (2014)
8	Lokot & Diakopoulos (2016); Sundar & Nass (2001)
7	Dörr & Hollnbuchner (2017); Hamilton & Turner (2009); Montal & Reich (2017)
6	Caswell & Dörr (2018); Gynnild (2014); Levy (2012); Lewis & Westlund (2015a); Lindén (2017); Meyer (1988); Sundar (1999); Sundar (2008)
5	Appelman & Sundar (2016); Broussard (2015); Cohen *et al.* (2011); Deuze (2005); Hovland & Walter (1951); Jung *et al.* (2017); Karlsen & Stavelin (2014); Lewis & Westlund (2015b); Pavlik (2000); Powers (2012); Reiter & Dale (2000); Stavelin (2013); Waddell (2018)

† Only citations mentioned 5 times or more are indicated.

Theoretical contributions published in an academic publication[4] were in great part concerned with exploring sociological aspects that relate to news making as well as to digital and algorithmic transformations (Anderson, 2012; Coddington, 2015; Deuze, 2005; Dörr & Hollnbuchner, 2017; Flew *et al.*, 2012; Gillespie, 2014; Gynnild, 2014; Lewis & Westlund, 2015a; Lewis & Westlund, 2015b; Napoli, 2014; Pavlik, 2000). That being said, the suggested sociological lenses some of these publications put forward remained largely unexploited in the articles I analysed; even if a handful of studies on the practice of automated journalism put these contributions at the centre of their analysis: (i.e., Anderson, 2012 and Napoli, 2014 in Dörr, 2016; Anderson, 2012 in Young & Hermida, 2015; Lewis & Westlund, 2015a in Thurman *et al.*, 2017; Deuze, 2005 in Lindén, 2017) none of them operationalised them as variables.

Out of all the theoretical contributions that were dealing with sociological aspects, only a handful advised to engage with well-established theoretical frameworks to look at how algorithms are transforming journalism. Dörr & Hollnbuchner (2017) recommended using traditional theories of ethics (i.e., deontology, utilitarianism, virtue ethics and contractualism) while Napoli (2014) pointed at institutional theory to emphasise how a social constructivist approach as well as a focus on the concept of Institutional isomorphism could help investigate algorithmic media consumption and production. Along with the same institutional lenses, Anderson (2012) recommended using Bourdieu's field theory to add "a vector of power dynamics" to the field of technological innovation, which he described (p. 1013) as being "too often understood from within an "all boats will rise" mentality." Although institutionalism is actively used – either partly or entirely – in a few studies on professionals (Dörr, 2016; Kim & Kim, 2017; Kim & Kim, 2018), the Bourdieusian lenses suggested by Anderson remained unexplored in the publications I systematically retrieved. Outside the corpus, though, Field theory is at the heart of Wu's *et al.* (2019a) of how algorithmic automation reshapes the journalistic field.

Finally, a last strand of studies visible through my listing of most-cited references are contributions focusing on readers' perceptions and evaluation of credibility, whether they focus on communication materials overall (Hovland & Weiss, 1951), newspapers (Meyer, 1988), printed and online news (Sundar, 1999), online news only (Sundar & Nass, 2001), technological aspects of digital media (Sundar, 2008) or message content (Appelman & Sundar, 2016). Contrarily to the sociological

[4] Other publications of this type included a conference paper (Hamilton & Turner, 2009) and a news article (Levy, 2012).

frameworks that I mentioned above, these psychologically-inspired studies – which can be empirical or theoretical – are largely operationalised as variables in the articles that investigated readers' perceptions (Sundar, 1999 in Clerwall, 2014, Haim & Graefe, 2017, Kim & Lee, 2019, Graefe *et al.,* 2018, Melin *et al.,* 2018 and Wu, 2020; Sundar, 2008 in Waddell, 2018, Waddell, 2019b and Wu, 2020; Sundar & Nass, 2001 in Zheng *et al.*, 2018; Meyer, 1988 in Liu & Wei, 2019, Wölker & Powell, 2018, Tandoc *et al.*, 2020 and Wu, 2020; Appelman & Sundar, 2016 in Waddell, 2018, Waddell, 2019a, Waddell, 2019b and Liu & Wei, 2019).

In sum, although studies that adopt psychological approaches to focus on readers' perceptions of automated journalism tend to fully engage with theory, the active use of sociological frameworks to investigate media labour remains largely unexploited. Of the multitude of theoretical contributions dealing with sociological aspects that have been mentioned, only a couple engage with well-established theories that can be employed to investigate automated journalism practices: Institutional theory, which is actively used in three studies (Dörr, 2016; Kim & Kim, 2017; Kim & Kim, 2018), and Field theory, which, so far, remains unused to study automated journalism only.

Conclusions

In this article, I analysed some of the features of a selection of scholarship on automated journalism in order to contribute a research agenda that advances guidelines and suggestions for future investigations into this area. First of all, I noticed that, although the expression "robot journalism" has largely been called into question, there is little debate as to whether "automated journalism" should be preferred, even if it could also refer to a wide range of automated tasks within newsrooms. Second, I became conscious of the fact that limiting myself to scholarship written in English prevents me from engaging with the full picture of automated journalism, and called for a review of scholarship published in other languages, such as in French, Spanish and Portuguese as they open a research window in Europe, the Americas and Africa, but also in local Asian languages, especially in order to better understand the strategies developed in a country like China. Third, I demonstrated the need to take a better look at individual and collective impacts the technology has on media practitioners and other associated actors to discern common patterns across media organisations, while studies on readers predominantly showed that they perceive human-written content and automated content in similar ways. Lastly, I found that, although psychologically-inspired frameworks were well exploited in studies focusing on readers' perceptions, sociological lenses were, in comparison, underused to look at automated journalism practices.

On a last note, while looking at the use of well-established theories to investigate automated journalism with a focus on media labour, I observed that Institutionalism was actively used in a couple of studies, but that the Bourdieusian lenses that Anderson suggested (2012) were missing. Therefore, the framework provided by field theory could be worth exploring in future studies exploring the individual and collective impacts the technology has on media labour, to find common patterns across media organisations.

Data availability

Underlying data

Zenodo: Supporting dataset – A systematic review of automated journalism scholarship: guidelines and suggestions for future research

https://doi.org/10.5281/zenodo.4442328 (Danzon-Chambaud, 2021a)

This project contains the following underlying data:

- Dataset_Systematic_Review_Automated_Journalism.xlsx (This dataset supports the following article, conditionally accepted for publication in Open Research Europe: "A systematic review of automated journalism scholarship: guidelines and suggestions for future research.")

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Reporting guidelines

Zenodo: PRISMA checklist for 'A systematic review of automated journalism scholarship: guidelines and suggestions for future research'.

https://doi.org/10.5281/zenodo.4444842 (Danzon-Chambaud, 2021b)

Zenodo: PRISMA flow diagram for 'A systematic review of automated journalism scholarship: guidelines and suggestions for future research'.

https://doi.org/10.5281/zenodo.4444916 (Danzon-Chambaud, 2021c)

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

References

Anderson CW: **Towards a sociology of computational and algorithmic journalism.** *New Media Soc.* 2012; **15**(7): 1005–1021.
Publisher Full Text

Appelman A, Sundar SS: **Measuring message credibility: Construction and validation of an exclusive scale.** *Journal Mass Commun Q.* 2016; **93**(1): 59–79.
Publisher Full Text

Ausserhofer J, Gutounig R, Oppermann M, *et al.*: **The datafication of data journalism scholarship: Focal points, methods, and research propositions for the investigation of data-intensive newswork.** *Journalism.* 2017; **21**(7): 950–973.
Publisher Full Text

Blankespoor E, deHaan E, Zhu C: **Capital market effects of media synthesis and dissemination: Evidence from robo-journalism.** *Review of Accounting Studies.* 2018; **23**: 1–36.
Publisher Full Text

Broussard M: **Artificial intelligence for investigative reporting.** *Digit Journal.* 2015; **3**(6): 814–831.
Publisher Full Text

Bucher T: **'Machines don't have instincts': Articulating the computational in journalism.** *New Media Soc.* 2017; **19**(6): 918–933.
Publisher Full Text

Carlson M: **The Robotic Reporter.** *Digit Journal.* 2015; **3**(3): 416–431.
Publisher Full Text

Carlson M: **Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism.** *New Media Soc.* 2018; **20**(5): 1755–1772.
Publisher Full Text

Carneiro dos Santos M: **Narrativas automatizadas e a geração de textos jornalísticos: A estrutura de organização do lead traduzida em código.** *Brazilian Journalism Research.* 2016; **12**(1): 160–185.
Publisher Full Text

Caswell D, Dörr K: **Automated Journalism 2.0: Event-driven narratives.** *Journalism Practice.* 2018; **12**(4): 477–496.
Publisher Full Text

Clerwall C: **Enter the robot journalist.** *Journalism Practice.* 2014; **8**(5): 519–531.
Publisher Full Text

Coddington M: **Clarifying journalism's quantitative turn.** *Digit Journal.* 2015; **3**(3): 331–348.
Publisher Full Text

Cohen S, Li C, Yang J, *et al.*: **Computational Journalism: A call to arms to database researchers.** *5th Biennial Conference on Innovative Data Systems Research (CIDR '11).* Asilomar, CA: January 9-12, 2011, 2011.
Reference Source

Danzon-Chambaud S: **Supporting dataset - A systematic review of automated journalism scholarship: guidelines and suggestions for future research [Data set].** *Zenodo.* 2021a.
http://www.doi.org/10.5281/zenodo.4442328

Danzon-Chambaud S: **PRISMA checklist for 'A systematic review of automated journalism scholarship: guidelines and suggestions for future research'.** *Zenodo.* 2021b.
http://www.doi.org/10.5281/zenodo.4444842

Danzon-Chambaud S: **PRISMA flow diagram for 'A systematic review of automated journalism scholarship: guidelines and suggestions for future research'.** *Zenodo.* 2021c.
http://www.doi.org/10.5281/zenodo.4444916

Deuze M: **What is journalism?: Professional identity and ideology of journalists reconsidered.** *Journalism.* 2005; **6**(4): 442–464.
Publisher Full Text

Diakopoulos N: **Algorithmic Accountability.** *Digit Journal.* 2015; **3**(3): 398–415.
Publisher Full Text

Diakopoulos N: **Automating the News: How Algorithms Are Rewriting the Media.** Cambridge, MA: Harvard University Press. 2019.
Reference Source

Díaz-Noci J: **Artificial intelligence systems-aided news and copyright: Assessing legal implications for journalism practices.** *Future Internet.* 2020; **12**(5): 85.
Publisher Full Text

Dierickx L: **Du statut d'objet technique à celui d'objet social dans le contexte de la production automatisée d'informations journalistiques.** *Terminal.* 2018; 122.
Publisher Full Text

Dierickx L: **Information automatisée et nouveaux acteurs des processus journalistiques.** *Sur le journalisme, About journalism, Sobre jornalismo.* 2019; **8**(2): 154–167.
Publisher Full Text

Dörr KN: **Mapping the field of Algorithmic Journalism.** *Digit Journal.* 2016; **4**(6): 700–722.
Publisher Full Text

Dörr KN, Hollnbuchner K: **Ethical challenges of algorithmic journalism.** *Digit Journal.* 2017; **5**(4): 404–419.
Publisher Full Text

Esteva A, Kuprel B, Novoa RA, *et al.*: **Dermatologist-level classification of skin cancer with deep neural networks.** *Nature.* 2017; **542**(7639): 115–118.
PubMed Abstract | **Publisher Full Text**

Falk Eriksen M: **When robots become colleagues: A case study of the Danish news agency.** Ritzau (Master thesis, Aarhus University and Hamburg University). Retrieved from personal communication, 2 January 2020, 2018.

Fanta A: **Putting Europe's robots on the map: Automated journalism in news agencies.** (Research paper, Reuters Institute for the Study of Journalism). 2017.
Reference Source

Flew T, Spurgeon C, Daniel A, *et al.*: **The promise of computational journalism.** *Journalism Practice.* 2012; **6**(2): 157–171.
Publisher Full Text

Ford H, Hutchinson J: **Newsbots that mediate journalist and audience relationships.** *Digit Journal.* 2019; **7**(8): 1013–1031.
Publisher Full Text

Gillespie T: **The relevance of algorithms.** In: Gillespie, T., P. J. Boczkowski, & K. A. Foot (Eds.) *Media Technologies: Essays on Communication, Materiality, and Society.* Cambridge, MA: MIT Press, 2014.
Reference Source

Graefe A: **Guide to Automated Journalism.** (Research paper, Tow Center for Digital Journalism). 2016.
Reference Source

Graefe A, Haim M, Haarmann B, *et al.*: **Readers' perception of computer-generated news: Credibility, expertise, and readability.** *Journalism.* 2018; **19**(5): 595–610.
Publisher Full Text

Graefe A, Bohlken N: **Automated journalism: A meta-analysis of readers' perceptions of human-written in comparison to automated news.** *Media and Communication.* 2020; **8**(3): 50–59.
Publisher Full Text

Graves L : **Understanding the promise and limits of automated fact-checking.** (Research paper, Reuters Institute for the Study of Journalism). 2018.
Reference Source

Gynnild A: **Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets.** *Journalism.* 2014; **15**(6): 713–730.
Publisher Full Text

Haim M, Graefe A: **Automated news.** *Digit Journal.* 2017; **5**(8): 1044–1059.
Publisher Full Text

Hamilton JT, Turner F: **Accountability through algorithm: Developing the field of computational journalism**. *Developing the Field of Computational Journalism.* Stanford, CA: July 27-31, 2009, 2009.
Reference Source

Hovland CI, Weiss W: **The influence of source credibility on communication effectiveness.** *Public Opin Q.* 1951; **15**(4): 635–650.
Publisher Full Text

Jung J, Song H, Kim Y, *et al.*: **Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists.** *Comput Human Behav.* 2017; **71**: 291–298.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Jones B, Jones R: **Public service chatbots: Automating conversation with BBC News.** *Digit Journal.* 2019; **7**(8): 1032–1053.
Publisher Full Text

Karlsen J, Stavelin E: **Computational journalism in Norwegian newsrooms.** *Journalism Practice.* 2014; **8**(1): 34–48.
Publisher Full Text

Katz DM, Bommarito 2nd MJ, Blackman J: **A general approach for predicting the behavior of the Supreme Court of the United States.** *PLoS One.* 2017; **12**(4): e0174698.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kim D, Kim S: **Newspaper companies' determinants in adopting robot journalism.** *Technol Forecast Soc Change.* 2017; **117**: 184–195.
Publisher Full Text

Kim D, Kim S: **Newspaper journalists' attitudes towards robot journalism.** *Telematics and Informatics.* 2018; **35**(2): 340–357.
Publisher Full Text

Kim D, Lee J: **Designing an algorithm-driven text generation system for personalized and interactive news reading.** *Int J Hum Comput Interact.* 2018; **35**(2): 109–122.
Publisher Full Text

Lemelshtrich Latar N: **The Robot Journalist in the Age of Social Physics: The End of Human Journalism?** In Einav, G. (Ed.) *The New World of Transitioned Media. The Economics of Information, Communication, and Entertainment*. Cham, CH: Springer, 2015; 65–80.
Publisher Full Text

Lemelshtrich Latar N: **Robot Journalism.** In Lemelshtrich Latar, Noam (Ed.) *Robot Journalism: Can Human Journalism Survive?* Singapore: World Scientific, 2018; 29–40.
Publisher Full Text

Levy S: **Can an algorithm write a better news story than a human reporter?** Wired. 2012.
Reference Source

Lewis SC, Sanders AK, Carmody C: **Libel by algorithm? Automated journalism and the threat of legal liability.** *Journal Mass Commun Q.* 2019; **96**(1): 60–81.
Publisher Full Text

Lewis SC, Westlund O: **Big Data and Journalism.** *Digit Journal.* 2015a; **3**(3): 447–466.
Publisher Full Text

Lewis SC, Westlund O: **Actors, actants, audiences, and activities in cross-media news work.** *Digit Journal.* 2015b; **3**(1): 19–37.
Publisher Full Text

Limia M: **Riesgos, desafíos e implicaciones éticas del periodismo robot.** Paper presented at the 2019 conference of the International Association for Media and Communication Research. Madrid, ES: July 7-11, 2019.

Lindén CG: **Decades of automation in the newsroom.** *Digit Journal.* 2017; **5**(2): 123–140.
Publisher Full Text

Lindén CG, Tuulonen H: **News automation: The rewards, risks and realities of 'machine journalism'.** Frankfurt, DE: WAN-IFRA. 2019.
Reference Source

Lindén CG, Dierickx L: **Robot Journalism: The damage done by a metaphor.** *Unmediated: Journal of Politics and Communication.* 2019; **2**: 152–155.
Reference Source

Liu B, Wei L: **Machine authorship *in situ*.** *Digit Journal.* 2019; **7**(5): 635–657.
Publisher Full Text

Lokot T, Diakopoulos N: **News bots.** *Digit Journal.* 2016; **4**(6): 682–699.
Publisher Full Text

Marconi F: **Newsmakers: Artificial Intelligence and the Future of Journalism.** New York, NY: Columbia University Press. 2020.
Publisher Full Text

Medvedeva M, Vols M, Wieling M: **Using machine learning to predict decisions of the European Court of Human Rights.** *Artif Intell Law.* 2020; **28**: 237–266.
Publisher Full Text

Melin M, Bäck A, Södergård C, *et al.*: **No Landslide for the human journalist - An empirical study of computer-generated election news in Finland.** *IEEE Access.* 2018; **6**: 43356–43367.
Publisher Full Text

Meyer P: **Defining and measuring credibility of newspapers: Developing an Index.** *Journal Mass Commun Q.* 1988; **65**(3): 567–574.
Publisher Full Text

Montal T, Reich Z: **I, Robot. You, Journalist. Who is the Author?** *Digit Journal.* 2017; **5**(7): 829–849.
Publisher Full Text

Monti M: **Automated journalism and freedom of information: ethical and juridical problems related to AI in the press field.** Opinio Juris in Comparatione 1. 2019.
Reference Source

Napoli PM: **Automated media: An institutional theory perspective on algorithmic media production and consumption.** *Communication Theory.* 2014; **24**(3): 340–360.
Publisher Full Text

Ojomo O, Ikem V: **Artificial intelligence (AI) and robot journalism adoption: ethical considerations from a developing country's perspective.** Paper presented at the 2019 conference of the International Association for Media and Communication Research. Madrid, ES: July 2019; 7–11.

Pavlik J: **The Impact of Technology on Journalism.** *Journal Stud.* 2000; **1**(2): 229–237.
Publisher Full Text

Plattner T, Orel D: **Addressing micro-audiences at scale.** Computation + Journalism Symposium 2019. Miami, FL: February 2019; **2019**: 1–2.

Powers M: **"In forms that are familiar and yet-to-be invented": American journalism and the discourse of technologically specific work.** *Journal of Communication Inquiry.* 2012; **36**(1): 24–43.
Publisher Full Text

Reiter E, Dale R: **Building Natural Language Generation Systems.** Cambridge, UK: Cambridge University Press. 2000.
Publisher Full Text

Rojas Torrijos: **Automated sports coverages. Case study of bot released by The Washington Post during Río 2016 and Pyeongchang 2018 Olympics.** *Revista Latina de Comunicación Social.* 2019; 1729–1747.
Publisher Full Text

Salawu A: **Robot journalism and the 4IR: Implications for practice and training.** Paper presented at the 2019 conference of the International Association for Media and Communication Research. Madrid, ES: July 2019; 2019; 7–11.

Sánchez Gonzáles HM, Sánchez Gonzáles M: **Los bots como servicio de noticias y de conectividad emocional con las audiencias: El caso de Politibot.** Doxa.Comunicación, 2017; **25**: 63–84.
Publisher Full Text

Stavelin E: **Computational journalism: When journalism meets programming.** (PhD thesis, University of Bergen). 2013.
Reference Source

Stray J: **Making artificial intelligence work for investigative journalism.** *Digit Journal.* 2019; **7**(8): 1076–1097.
Publisher Full Text

Sundar SS: **Exploring Receivers' Criteria for Perception of Print and Online News.** *Journal Mass Commun Q.* 1999; **76**(2): 373–386.
Publisher Full Text

Sundar SS, Nass C: **Conceptualizing sources in online news.** *Journal of Communication.* 2001; **51**: 52–72.
Publisher Full Text

Sundar SS: **The MAIN Model: A heuristic approach to understanding technology effects on credibility.** In Metzger, M. J. & Flanagin, A. J. (Eds.) *Digital Media, Youth, and Credibility*. The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning. Cambridge, MA: MIT Press, 2008; 73–100.
Reference Source

Tandoc EC, Yao LJ, Wu S: **Man vs. Machine? The impact of algorithm authorship on news credibility.** *Digit Journal.* 2020; **8**(4): 548–562.
Publisher Full Text

Thurman N, Dörr K, Kunert J: **When reporters get hands-on with robo-writing.** *Digit Journal.* 2017; **5**(10): 1240–1259.
Publisher Full Text

Túñez-López JM, Toural-Bran C, Cacheiro-Requeijo S: **Uso de bots y algoritmos para automatizar la redacción de noticias: percepción y actitudes de los periodistas en España.** *Profesional de la Información.* 2018; **27**(4): 750–758.
Publisher Full Text

Túñez-Lopez JM, Toural-Bran C, Valdiviezo-Abad C: **Automation, bots and algorithms in newsmaking: Impact and quality of artificial journalism.** *Revista Latina de Comunicación Social.* 2019; **74**: 1411–1433.
Publisher Full Text

van Dalen A: **The algorithms behind the headlines.** *Journalism Practice.* 2012; **6**(5–6): 648–658.
Publisher Full Text

van der Kaa H, Krahmer E: **Journalist versus news consumer: The perceived credibility of machine written news.** Computation + Journalism Symposium 2014. New York, NY: October 24-25, 2014, 2014.
Reference Source

Visvam Devadoss AK, Thirulokachander VR, Devadoss AKV: **Efficient daily news platform generation using natural language processing.** *Int J Inf Technol.* 2019; **11**: 295–311.
Publisher Full Text

Young ML, Hermida A: **From Mr. and Mrs. Outlier to Central Tendencies.** *Digit Journal.* 2015; **3**(3): 381–397.
Publisher Full Text

Waddell TF: **A robot wrote this?** *Digit Journal.* 2018; **6**(2): 236–255.
Publisher Full Text

Waddell TF: **Attribution practices for the man-machine marriage: How perceived human intervention, automation metaphors, and byline location affect the perceived bias and credibility of purportedly automated content.** *Journalism Practice.* 2019a; **13**(10): 1255–1272.
Publisher Full Text

Waddell TF: **Can an algorithm reduce the perceived bias of news? Testing the effect of machine attribution on news readers' evaluations of bias, anthropomorphism, and credibility.** *Journal Mass Commun Q.* 2019b; **96**(1): 82–100.
Publisher Full Text

Weeks L: **Media law and copyright implications of automated journalism.** *Journal of Intellectual Property and Entertainment Law.* 2014; **4**(1): 67–94.
Reference Source

Wölker A, Powell TE: **Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism.** *Journalism.* 2018.
Publisher Full Text

Wu S, Tandoc EC, Salmon CT: **A Field analysis of journalism in the automation age: Understanding journalistic transformations and struggles through structure and agency.** *Digit Journal.* 2019a; **7**(4): 428–446.
Publisher Full Text

Wu S, Tandoc EC Jr, Salmon CT: **Journalism Reconfigured.** *Journal Stud.* 2019b; **20**(10): 1440–1457.
Publisher Full Text

Wu Y: **Is automated journalistic writing less biased? An experimental test of auto-written and human-written news stories.** *Journalism Practice.* 2020; **14**(8): 1008–1028.
Publisher Full Text

Zheng Y, Zhong B, Yang F: **When algorithms meet journalism: The user perception to automated news in a cross-cultural context.** *Comput Human Behav.* 2018; **86**: 266–275.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 23 July 2021

https://doi.org/10.21956/openreseurope.14166.r27043

© **2021 Kunert J.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

✔ **Jessica Kunert**
Institute for Journalism and Communication Studies, University of Hamburg, Hamburg, Germany

I would like to thank Samuel Danzon-Chambaud for this thorough review of automated journalism scholarship. He evaluates semantical, chronological and geographical features of the articles and discusses their implications for the future.

I have a few comments. First, to make the argument for a review of this scholarship even stronger, I suggest adding a few more recent examples of the application in the introduction. Germany, for example, is now using it for election coverage (MDR; federal state elections 2021 in Sachsen-Anhalt) and for sports reporting beyond amateur football (BR; basketball). Not that I personally need that hint that this field of research is crucial (please read this with a wink), but it would give people who are not in the field a better notion of what is going on and why studying automated journalism is important.

My second, more major suggestion is to add more structure to the article to make it more accessible. While I do think the article is a good read, I'd prefer to see it having information that can be digested (and then cited) at first glance. I feel that in the introduction, when talking about automated journalism being a part of computational studies, there could be more clarity on where the individual research foci are situated. Moreover, one research field is missing, namely the effect on journalistic roles in conjunction with the affordances of automated journalism. Let me shamelessly bring my own work forward here – the Thurman, Dörr, & Kunert 2017 article deals with this, and more recently my 2020 article on automated sports journalism.

Adding more structure could also mean adding more tables instead of really long sentences. This for example could apply to putting the research aims of 'practice' into tables for a better overview (p. 6 to 8). The same goes for the multitude of theories used (p. 9).

As for the methods, I agree that wading through the scholarship is hard, especially because so many articles do not concern journalism but are rather tech-oriented. Samuel Danzon-Chambaud also says himself later on that focusing only on articles published in English might be a problem – but I don't see this as a weakness of this article, but rather as a call to action to band together and

get an overview of scholarship around the world in international teams. However, I would have liked to see the codebook in full; maybe it's possible to upload it.

The results are quite exciting, especially the discussion of the terms. My suggestion would be to add a paragraph on the 'historical' use of the terms – 'robot journalism' was an early term (quite possibly because there simply wasn't a better one, at least not such a colourful one), and things are changing rapidly. The latest one I've heard is 'augmented journalism', referring to the human-machine partnership. Maybe this could be added to give a look into the future? Moreover, I think a key finding is that sociological lenses are under exploited – researchers should definitely read up on theory in this field (but of course, I can also speak for myself).

One more thing on presentation: It would help immensely if the figures were labelled more clearly, meaning that for example in figure 2, it's hard to see at first glance just how many studies used content analysis and so on. Adding the n to the individual categories would help the matter.

In sum, this a valuable article that demonstrates where research has made headway – and where more work needs to be done.

References

1. Kunert J: Automation in Sports Reporting: Strategies of Data Providers, Software Providers, and Media Outlets. *Media and Communication*. 2020; **8** (3): 5-15 Publisher Full Text

Are the rationale for, and objectives of, the Systematic Review clearly stated?

Yes

Are sufficient details of the methods and analysis provided to allow replication by others?

Partly

Is the statistical analysis and its interpretation appropriate?

Yes

Are the conclusions drawn adequately supported by the results presented in the review?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Automated journalism; computational journalism

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 27 April 2021

https://doi.org/10.21956/openreseurope.14166.r26541

© 2021 Porlezza C. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Colin Porlezza
Institute of Media and Journalism, Università della Svizzera italiana, Lugano, Switzerland

In his article, Danzon-Chambaud offers a thorough overview of the latest empirical research on automated journalism published in peer-reviewed journals. The rationale for looking into this particular topic relates to two aspects: first, AI and automated journalism have become quite common in (elite) news organizations around the world. Second, as a consequence, this phenomenon has come under close scrutiny of journalism studies as well.

The contribution carried out a systematic literature review by selecting articles based on empirical research on automated journalism. The articles had to be specifically about automated journalism and not generally about AI and journalism, which means that only articles dealing either with "the process of using software or algorithms to automatically generate news stories without human intervention – after the initial programming of the algorithm" (Graefe, 2016, p. 14)[1], or the auto generation of journalistic texts were taken into account. In a second step, the author carried out a descriptive analysis of the specific terms used when referring to automated journalism, the chronological and geographical traits, the fields of inquiry and the methods used in the scholarship, and whether and which theories the articles were using.

The article offers a useful and thorough insight into a quickly growing field of enquiry and is therefore a good starting point if scholars want to familiarize themselves - not necessarily with the findings of the empirical studies - but with the way digital journalism studies tried to tackle the issue of automated journalism. It analyzes therefore the characteristics of research into automated journalism to identify dominant patterns and themes regarding semantics, methods and theories. This article is a first of its kind, which makes for a relevant contribution because - as the article demonstrates - the theoretical approaches used in the investigations are limited.

I would like to focus on two specific issues: first, the semantic question whether the expression automated journalism is too narrow and does not reflect media practitioners' perspectives. And second, a discussion about what kind of theories to use in the future, and why.

Let's start with the first question: the author is certainly right in stating that automated journalism has been used in a rather narrowish way. By looking at the publications in the field it becomes quite clear that automated journalism is the preferred "term within industry and academia" (Danzon-Chambaud 2021). However, there has not been a coherent use and application of the different terms in the field, and Wu's *et al*. (2019, p. 1453)[2] observation that the expression "does not reflect media practitioners' views" is equally true. The main reason for this incoherence - we could also characterize it in a more positive way as a plurality - shows first of all a high vitality in the field. As Feola (2015)[3] states, this is quite common in the social sciences, where different paradigms and concepts coexist at a time. Particularly in a new and emerging field of research such as this one, it might also be argued that a plurality of expressions "with 'loose' meanings that also benefit from a strong metaphoric power can be highly effective to stimulate research and action, and to create a much needed common ground for scholars from different disciplines, as

well as among scientists and stakeholders, which facilitates inter- and transdisciplinarity" (Feola, 2015). A complex field such as the one revolving around automated journalism is characterized by increasing interdisciplinarity, and therefore needs not only to assess the conceptual diversity, but also the different concepts' blind spots as well as the different research approaches. The challenge therefore is to find a useful and creative way "to foster a dialog around the potential and complementarities of different concepts: what concepts can be applied to study different types of systems, what prescriptive assumptions inform them, what concepts connect with what research paradigm (...)" (idem). What is needed - and in this I convene with the Danzon-Chambaud's critique - is a more structured engagement with the different concepts, but also a critical discussion that includes different disciplines and traditions of thought. A good example of such a reflective contribution can be seen in Lewis, Guzman and Schmidt's (2019)[4] paper on "Automation, Journalism, and Human–Machine Communication", an article suggesting that "research focused on automated journalism, has much to learn from Human-Machine Communication (HMC)" - which is a theoretical paper and therefore not included in the author's review. Hence, for a more complete overview of the field's engagement with the phenomenon of automated journalism, it might have proven helpful to include theoretical contributions in the review as well.

This brings me to the second point of my review, that is, the question of what kind of theories to use in the future. At the very end of his article, the author states "while looking at the use of well-established theories to investigate automated journalism with a focus on media labour, I observed that Institutionalism was actively used in a couple of studies, but that the Bourdieusian lenses that Anderson suggested (2012) were missing. Therefore, the framework provided by field theory could be worth exploring in future studies exploring the individual and collective impacts the technology has on media labour, to find common patterns across media organisations" (Danyon-Chambaud 2021). While the author is certainly right in stating that a field theory perspective has not been used frequently, it is nevertheless surprising that - given the increasingly interdisciplinary character of the field - there are no references to other promising theoretical frameworks such as Human-Machine-Communication (HMC), Actor-Network-Theory (ANT) or Science and Technology Studies (STS) precisely when it comes to future endeavors in the field. Especially when it comes to questions of power shifts between humans and technology, field theory is not the only framework able to "add an additional vector", but HMC for instance is equally suited to investigate the social positioning of technology by looking at "how a person interprets what a particular technology is in relation to themselves, the factors contributing to such interpretations, and, in turn, how such conceptualizations inform their interactions" (Guzman & Lewis, 2019)[5]. If the author chooses to point out a Bordieuan lens, it would have been useful to offer more room to the author's explanations given the many theoretical alternatives.

All in all, the article is a helpful snapshot of how the field of digital journalism studies engages with the emergent phenomenon of automated journalism. The findings could also be used to initiate a wider discussion on the current plurality of concepts and on how to deal with the differences, complementarities and blind spots of the theoretical frameworks, particularly as the field becomes more interdisciplinary.

References

1. Graefe A: Guide to Automated Journalism. *Tow Center for Digital Journalism, Columbia University*. 2016.
2. Wu S, Tandoc E, Salmon C: A Field Analysis of Journalism in the Automation Age: Understanding Journalistic Transformations and Struggles Through Structure and Agency. *Digital Journalism*. 2019;

7 (4): 428-446 Publisher Full Text
3. Feola G: Societal transformation in response to global environmental change: A review of emerging concepts.*Ambio*. 2015; **44** (5): 376-90 PubMed Abstract | Publisher Full Text
4. Lewis S, Guzman A, Schmidt T: Automation, Journalism, and Human–Machine Communication: Rethinking Roles and Relationships of Humans and Machines in News. *Digital Journalism*. 2019; **7** (4): 409-427 Publisher Full Text
5. Guzman A, Lewis S: Artificial intelligence and communication: A Human–Machine Communication research agenda. *New Media & Society*. 2020; **22** (1): 70-86 Publisher Full Text

Are the rationale for, and objectives of, the Systematic Review clearly stated?
Yes

Are sufficient details of the methods and analysis provided to allow replication by others?
Yes

Is the statistical analysis and its interpretation appropriate?
Yes

Are the conclusions drawn adequately supported by the results presented in the review?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Digital journalism studies; artificial intelligence; media accountability and transparency; journalism innovation.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.