Automated news in practice:
Changing the journalistic *doxa* during COVID-19, at the BBC and across media organisations

Samuel Danzon-Chambaud, MA, MJ

Thesis submitted for the award of PhD

School of Communications
Dublin City University

Supervisory Panel:

Dr. Alessio Cornia
Prof. Kevin Rafter
Dr. David Robbins

January 2023

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of ....PhD.......... is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work. Signed:

_____ ID No.:_____18213855_____ Date: ___10/01/2023___

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANT: Actor-network theory
API: Application programming interface
CMS: Content management system
NLG: Natural language generation
SEO: Search engine optimisation

ABSTRACT

AUTOMATED NEWS IN PRACTICE: CHANGING THE JOURNALISTIC *DOXA* DURING COVID-19, AT THE BBC AND ACROSS MEDIA ORGANISATIONS

Samuel Danzon-Chambaud

This PhD thesis explores the deployment of automated text generation for journalistic purposes—also known as automated news or automated journalism—within newsrooms. To evaluate its perceived impacts on the work of media practitioners, I rely on Bourdieu's Field theory, but also make use of Actor-network theory to detail its adoption at a more descriptive level. This study is based on various case studies and on a mixed-methods framework that is essentially made of 30 semi-structured interviews conducted with media practitioners, technologists and executives working at 23 news organisations in Europe, North America and Australia; it also involves elements of a netnography as online material and screenshots were analysed as part of this process.

My empirical work starts with a descriptive account that includes three case studies: one on the use of automated news to cover COVID-19, another one on BBC's experiments with the technology and a last one that shows a cross-national comparison between three media types (i.e., public service media, news agencies and newspapers). I then move on to a more interpretative part where I examine media practitioners' reactions to automated news, analysing the challenges of having to rely on external datasets, the importance of acquiring a computational thinking mindset and tensions within and outside the field of journalism for this.

My research shows that the use of automated news implies structural changes to journalism practice and cannot be seen as a mere "tool of the trade". For practitioners, the most challenging part lies with being able to master both the uniqueness of journalistic work and a type of abstract reasoning close to computer programming. However, this could leave some being unable to adapt to this new computational spirit, which seems to be gradually taking root within newsrooms. As for future development of automated news systems, it remains to be seen if media organisations or platforms will have the upper hand in remaining at the centre of it.

1   INTRODUCTION

This PhD dissertation explores the deployment of automated text generation for journalistic purposes—also known as automated news or automated journalism—within newsrooms. To do this with a critical eye, I mostly rely on French sociologist Pierre Bourdieu's Field theory to evaluate the perceived impacts the technology has on the work of media practitioners, but also make use of Actor-network theory at a more descriptive level. The material for this thesis is mostly made of 30 semi-structured interviews conducted with media practitioners, technologists and executives working at 23 news organisations; it also involves elements of a netnography as online material and screenshots were examined in complement to these. They cover three case studies: one single case study on the use of automated news to cover COVID-19 and another one on how the BBC experimented with the technology, and one multiple case study that looks at its utilisation across three media types (i.e., public service media, news agencies, newspapers) and media systems, using Hallin and Mancini's typology (2004) as a framework of reference. My empirical results are then analysed to see how they connect to Bourdieu's Field theory and Actor-network theory, so as to come up with a critical outlook that lays out the opportunities as well as the more adverse effects induced by the technology. In this Introduction, I will first question the assumption that journalism has always been shaped by technology, then look into more nuanced perspectives of the relationship between the two. I will also give an overview of the use of algorithms in news production, and highlight the importance taken by automated news—which calls for better investigating its perceived impacts on the work of practitioners. Finally, I will detail my research thesis' outline.

1.1   JOURNALISM ALWAYS SHAPED BY TECHNOLOGY?

When studying the relationship between journalism and technology, it may be assumed that changes in journalism practice can be explained by looking at technological developments alone. Looking at the rapid transformations that

followed the rise of the Internet, Pavlik (2000, p. 299) argued that "journalism has always been shaped by technology". To him, the overall quality of journalistic work was to be improved with reporters being able to access online records on a deadline, and the use of the inverted pyramid—a journalistic form of writing that puts the most important information at the top of the story—was to be gradually abandoned for new multimedia elements such as 360-degree videos.

In hindsight, those predictions seem to be a little at odds with what is really happening today. Although reporters gained easier access to online resources, the heavy reliance on informational material, the increasingly desk-bound nature of journalistic work, ripping off news content published elsewhere without attribution (i.e., news "cannibalisation") and, overall, a more passive approach to news making do flourish in an online environment (Wheatley, 2020; Paulussen, 2012; Phillips, 2010; O'Neill & O'Connor, 2008). Besides, the inverted pyramid has turned out to be well-adapted to writing for online, as its emphasis on the '5Ws' (i.e., What, Why, When, Where, and Who) at the beginning of the story can match users' queries on search engines. This shows, if anything, that the relationship between journalism and technology is more complex than the simple assumption that "journalism has always been shaped by technology", and thus requires more nuanced perspectives.

Pavlik's thought, indeed, seems to be rooted in a *technological deterministic* perspective whereby technology follows a path on its own and has individuals, organisations, societies, cultures or economies adapt to it rather than the other way round (Regan Shade, 2003; Volti, 2020). In media and communication studies, examples of technological determinism can be found in Innis' (1950) *time* and *space* biases theory (Humphreys, 2010; Mateus, 2020), in which *time-biased* mediums—set to be immutable in nature and to last in time (e.g., message carved in stone)—give rise to societies that place much value in rituals and face-to-face interactions and favour stability and tradition, whereas *space-biased* mediums that can be easily be moved around but that are more ephemeral (e.g., print press, electronic media) foster social changes, secularism and expansion. Another example of technological determinism is McLuhan's claim that "the medium is the message" (1964), meaning that different mediums—be it television, print, or radio—can trigger different types of heuristics among audiences even though content remains the same: this can be

viewed as "a quintessential example of technological determinism" (Humphreys, 2010, p. 871) as communication technologies have, so to speak, taken a life of their own and supersede the message content creators intend to deliver.

Technological deterministic positions usually entail optimistic accounts of what the future holds (Regan Shade, 2003), as seen for instance in McLuhan's (1962, 1964) vision of a *global village* where electronic media—as they shorten communication lags—were to recreate small-town interactions at a global level (Howard, 2003). Similarly, the *information society* that gradually emerged as a continuum to the agrarian and industrial ages (see Bell, 1973)—and among which figures the *networked society* (Castells, 1996) characterised by endless flows that took over long-held boundaries—may have been perceived, for a time, as a way to end conflicts (Muir, 2020). That being said, technological deterministic views also encompass dystopian scenarios where technology gets "out-of-control" (Volti, 2020), just like the information society may result (Muir, 2020) in "the hegemony of faceless powers that remain unaccountable for their actions".

Although technological determinism was the dominant paradigm up until the mid-1980s (Regan Shade, 2003), it has since given way to a *social constructivist* or *social shaping* point of view which posits that control remains in the hands of underlying forces that lie behind the emergence or adoption of any given technology (Humphreys, 2010). For instance, the widespread use of smartphones may be considered the result of commercial ventures seeking to create integrated circuits that fulfil the government's requirements for space and defence programs, government contracts that supported the development of a Global Positioning System (GPS) and economic considerations that saw the manufacturing of such phones being outsourced to Asian countries where a cheaper labour force can be found (Volti, 2020). Among these social constructivist perspectives figures the *social construction of technology* or *SCOT* framework, which advances that social groups also have a say in the way a technological artefact is implemented after its original conception: an object is viewed and holds different interpretations depending on what type of users manipulate it, until a prevailing usage is reached (Bijker, 1995; 2008). Softer versions of technological determinism that acknowledge the role of external forces, but also stress that technology itself does have an important part to

play in its development, constitute yet another contribution to the field of science and technology studies (Adler, 2008; Humphreys, 2010), just like the idea that technology diffusion could take a social constructivist form first—where forces and actors actively participate in its shaping—and then becomes more determinate over time (Hughes, 1994; Adler, 2008; Volti, 2020).

Going opposite to Pavlik's "hard" technological deterministic view, we may rather see a social constructivist or soft technological deterministic connection between journalism and technology, starting with a mutual shaping relationship (see Boczkowski, 2004) between the two: if it is true that technological innovation does prompt changes, it is also accurate to say that these changes depend on how actors interpret, use, resist or adhere to them. In the section below, I will give an example of such a mutual shaping relationship by looking at the rise of American newspapers, then further stress the role that social context takes in the development of technologies shaping journalism practice. Finally, I will detail three different ways media practitioners may react to technological change within the newsroom.

1.1.1   Journalism and technology: a more complex relationship

A mutual shaping relationship where both journalism and technology intersect can be found in Schudson's account of the history of American newspapers. Schudson (1978) explains how the breakthroughs in printing technology that were made during the Industrial Revolution (e.g., all-iron and cylinder printing press, use of steam power)—along with the digging of canals and construction of railroads that made carrying this heavy equipment possible—did account for the circulation of cheaper mass newspapers. At the same time, Schudson also sheds light on how newspapers, including the new "penny press"[1] that came as a result of these developments, were in fact a driving force behind the demand for these types of technology. As an example of this, he mentions Frederick Koenig, the inventor of the steam-powered cylinder press, whose work was being subsidised by John Walter, the proprietor of the London

---

[1] Named as such because it sold for one penny a copy (as opposed to more formal newspapers that sold for six cents a copy), the American "penny press" democratised access to news titles between the 1830s and the 1860s, with a great number of copies being printed (Beasley, 2009). It featured sensational stories and a more casual style of writing. The *New York Sun,* the *New York Herald* and the *New York Tribune* figure among the main "penny press" titles of this time.

Times. Likewise, in the United States, penny press papers were always first in installing new printing machinery. In Schudson's own words (p. 33), "it may be more accurate to say that the penny press introduced steam power to American journalism than to say that steam brought forth the penny press".

In addition to Schudson's case for a mutual shaping relationship, Örnebring (2010) argues that the use of technology can be influenced by societal context, which results in turn in shaping journalism practice. To illustrate this, Örnebring (p. 65) stresses how the "discourse of speed", which is rooted in the capitalistic ideal of leveraging technology to increase production, has gradually become a "wholly naturalized element of journalism", giving rise for instance to the 24/7 news cycle. Similarly, the importance given to brevity—which relates to the same impetus for speed—led to adopting the inverted pyramid, which also revealed to be cost-efficient as telegraphing fees were based on word count. Most contemporary journalism scholarship, especially those grounded in social constructivism and inspired by science and technology studies, share similar views that stress that technology is as much influenced by social structures than by its material aspects only (Zamith & Braun, 2019).

Whether the relationship between journalism and technology is the product of a mutual shaping relationship or is entrenched within a wider value system, the way media practitioners envision this relationship is also important to take into account. In his analysis of discourses about "technologically specific forms of work"[2] in American journalism from 1975 to 2011, Powers (2012) sees three ways journalists react to new technological capacities being brought into the newsroom: first, by considering them an extension of existing occupational practices and values, which then triggers conversations on how to best harness them in order to enhance journalistic autonomy; second, by seeing them as a threat that needs to be "subordinated" because they do not correspond to occupational norms, which generally prompts a call to go back to core occupational practices and values and making those new forms of work look foreign, unnecessary and even dangerous;

---

[2] Powers defines these forms of work as being linked to the use of technology that eventually resulted in making journalistic claims. For instance, this could be about the way photography gave rise to photojournalism.

third, by evaluating whether these new forms of work can serve as a basis for a reinvention of occupational norms, even if what lies ahead is still unclear.

1.1.2 From wirephoto and "radio cars" to algorithms' use

Looking at the deployment of new technologies over time, we can see how they correspond to each of Powers' categories, starting with an extension of existing occupational practices and values. Taking greater access to telephone lines and the introduction of "radio cars" (i.e., cars equipped with radio communication devices) in 1930s' American newsrooms as an example, Mari (2018) observed that journalism practice was indeed enhanced through reporters being able to better reach out to their sources and gaining extra mobility on the ground[3]. In contrast, Zelizer (1995) illustrated how wirephoto, which came out at more or less the same time, was perceived by American journalists as a threat that needed to be subdued: they either depreciated the value that photography brought to the journalistic trade—for instance while insisting that text should take precedence over pictures—argued that themselves could act as professional photographers or described wirephoto as a "necessary evil" to meet audiences' expectations of high realism. Finally, regarding the reinvention of occupational norms, Boyles and Meisinger (2020) showed that, although newsroom librarians were among the first to be impacted by the introduction of digital technologies within American newspapers, they were also keen on seeking out new tasks like those left over by staffers who were made redundant— or managing book clubs.

　　An important development to be analysed here—to see how it connects to Powers' categories—is the emergence of computer-assisted reporting or "CAR" from the end of the 1960s to the mid-1990s, which preceded present-day computational journalism practices (to be discussed in section 2.2.3). Although Mari (2019) locates an earlier example of a data-driven story with the *Philadelphia Inquirer*, the computer-assisted reporting movement really began in the 1960s with investigations such as Philip Meyer's deep look into the Detroit riots in 1967, which showed that rioters were

---

[3] Mari also noted that greater access to a phone line came along with an increased workload for reporters, and that the introduction of radio cars meant that editors could better keep track of where they were, this on a continuous basis.

not so much less educated nor disadvantaged as they were frustrated with a lack of advancement, when compared with their White college graduate peers (Parasie, 2022): conducting such an investigation demanded quite a considerable amount of resources at the time, as shown in Meyer's use of survey results which were turned into punch cards so that they could be read by a then hefty mainframe computer (his team also included a political scientist, a statistician, a psychologist and a programmer). Eventually, these efforts helped his newspaper—the *Detroit Free Press*—earn a Pulitzer Prize. Meyer's work set the stage for other exemplary pre-data journalism stories in the 1970s and 1980s (Anderson, 2018; Parasie, 2022), even though these were still faced with fundamental limitations: first, for a time digitally stored information was hardly available to news outlets; second, the use of bulky mainframe computers was particularly cumbersome and—as described above—required many resources (Parasie, 2022). The introduction of personal computers or "PCs" around the beginning of the 1990s did not solve the issue of access to digitalised records (and associated costs), but this was partially addressed by journalist Elliott Jaspin's efforts in making mainframe computers' magnetic tapes readable on these new devices (Mari, 2019; Parasie, 2022).

In parallel with all these developments, early data journalism practices gradually became institutionalised, starting with Meyer's book *Precision Journalism* (1973), which advocated for training journalists in the basics of quantitative social science—or to put it in his words to do "social science in a hurry" (*ibid.*, p. 14). According to Anderson (2018), this book came as a result of social and political turmoil in the 1960s and early 1970s (e.g., John F. Kennedy, Malcolm X., Martin Luther King Jr. and Bobby Kennedy assassinations, the Detroit riots, the Summer of Love, the Watergate scandal, etc.), which saw journalistic craft evolve in one of two ways: either by becoming a more narrative and literary exercise as in the work of Tom Wolfe, Truman Capote and Hunter S. Thompson, or by taking a more quantitative-oriented and objective turn as defended by Meyer. Other than Meyer's book, the opening of the now National Institute for Computer-Assisted Reporting (i.e., NICAR)—which was founded by Jaspin at the turn of the 1990s—also helped further legitimise data journalism practices through establishing guidelines and providing training (Mari, 2022; Parasie; 2022), even though these were increasingly being perceived as elitist and high-level (Anderson, 2018). In the end, the history of CAR

shows that the forefathers of data journalism found "an original way to align technology with the core values of the profession" (Parasie, 2022, p. 69), thus corresponding to Powers' category of technological change as an extension of core occupational practices and values[4].

With the rise of the Internet in the mid-1990s and its quasi-ubiquitous status now, much attention has since been given to a specific aspect of computers' functioning, namely the use of algorithms, a step-by-step computerised command chain which has brought, in turn, new technologically specific forms of work within the newsroom. In general, algorithms tend to be discussed in the context of automated content curation (e.g., selecting news items to appear on a newsfeed, ranking and displaying search results), especially as there is a need to uncover "the warm human and institutional choices that lie behind these cold mechanisms" (Gillespie, 2014, p. 169). This stands even more true as algorithms play a crucial role in determining the type of information that is considered most relevant to users, thus impacting civic life. Consequently, it is then important to examine discourses that deal with the "cultural prominence of the notion of the algorithm, what this stands for, what it does and what it might reveal" (Beer, 2017, p. 11). Lastly, algorithms are also being seen as nurturing "filter bubbles", a social representation of the world that corresponds to a web user's own ideological beliefs (Pariser, 2011), which has prompted fears that it could create "echo chambers" further entrenching political divides (Sunstein, 2017). However, little evidence comes to support this—except for a minority of people who already have strong views—and other findings suggest that exposure to diverging views has even been on the rise (Fletcher & Jenkins, 2019).

Algorithms also have critical implications for journalism practice: first, they allow for the deployment of web metrics within newsrooms, potentially reducing journalists' autonomy as they have to factor readers' preferences into their own news judgement (Anderson, 2011); second, they can be programmed so that they directly contribute to news making. Diakopoulos (2019) shows that algorithms can be used this way to assist journalists with fact-checking or investigative pieces, or to generate

---

[4] Parasie (2022) also details how the use of computation led to a new form of "rankings" journalism (e.g., hospitals, schools, universities), which is often criticised and does not necessarily match the higher ideals of early data journalists.

content through automated news or "newsbots" on social media. In this respect, using algorithms for news production can be seen as one of the latest "technologically specific forms of work" described by Powers. Whether practitioners evaluate those as a continuation of existing norms, a threat to occupational values or an opportunity for reinvention, though, remains quite unclear. Wu, Tandoc and Salmon (2019a) found that news workers assume they have control over algorithmic news production[5], thus fitting into Powers' category of a continuation of existing norms. However, in another article (2019b), they underline that the very technical skills that are needed to be able to handle automation are met with resistance by senior journalists used to the traditional ways of doing journalism, or by practitioners who see those as being rather a programmer's job, thus making them look foreign or unnecessary as in Powers' second category. Finally, Milosavljević and Vobič (2021) illustrated that newsroom managers were holding a somewhat mixed discourse of "algorithmic sublime" that tend to mitigate the damaging effects these could have on journalism practice—such as leaving humans aside—to focus instead on the opportunity to augment journalism's public spirit.

1.2   ALGORITHMS IN NEWS PRODUCTION: MAIN APPLICATIONS

To get a better grasp on the use of algorithms for news production, I will give here a brief overview of their main domains of application: first, I will touch on the use of data mining techniques for investigative journalism, then look into the deployment of automated fact-checking to verify content at scale; second, I will delve into the rise of automated journalism and illustrate how it constitutes the most advanced application of algorithmic news production at the moment.

1.2.1   Data mining for investigative journalism and automated fact-checking

Looking, first, at data mining techniques, Diakopoulos (2019) illustrates how advanced machine learning models—in other words the use of algorithms to make

---

[5] In this study, Wu, Tandoc and Salmon focus on algorithmic automation, which news workers describe as incorporating a wide range of tasks such as data scraping for financial news, aggregating user-generated content and auto-publishing machine-written news.

statistical inferences or classifications based on a large corpus of data—were used in investigative journalism to retrieve newsworthy material off a massive amount of documents. He cites the work of *The Atlanta Journal Constitution*, which managed to expose 2,400 doctors that have been disciplined for sexual misconduct in the United States while examining more than 100,000 records, using for that an algorithm that scored and sorted through documents based on the likelihood that an abuse had, indeed, actually occurred.

That said, Stray (2019) highlighted how the use of such models in investigative journalism also comes with its own set of issues. He stressed that exposure to potential lawsuits in investigative journalism necessitates near perfect accuracy in the data mining techniques used. Stray detailed, in fact, how *The Los Angeles Times* accounted for the error margin of the machine learning model it used when examining whether 400,000 incident reports have been misclassified by the Los Angeles Police Department: even though the model was trained on 20,000 incidents collected over a year, the newsroom found a 24% error rate when reviewing a sample of 2,400 cases. The team then chose to feature estimates that compensate for this error margin. Stray also mentions difficulties in accessing the data in the first place, or the high-cost of deploying such models just for a one-off project. However, collaborative efforts that are being made in investigative journalism today could help solve these issues, as demonstrated by the growing use of machine learning techniques by the International Consortium of Investigative Journalists, in investigations like the "Implant Files", the "Mauritius Leaks", the "Luanda Leaks" or the "Pandora Papers" (Walker Guevara, 2019; Woodman, 2019; Díaz-Struck, Romera & Ledésert, 2020; Díaz-Struck *et al.*, 2021).

Another manifestation of algorithmic news production can be found in a range of initiatives aiming at automating fact-checking[6]. Diakopoulos (2019) identifies similar machine learning methods where a trained algorithmic model can be deployed on textual data so as to reveal claims that are worth fact-checking. This is for instance what Duke University Reporters' Lab has been doing to verify some of CNN's transcripts with a piece of software called "ClaimBuster". Diakopoulos also writes

---

[6] This part is based on a news article published with Poynter, which looked at a range of automated fact-checking initiatives (see Danzon-Chambaud, 2020).

about more basic methods that consist in matching textual data against a database made of previous fact-checks and of reliable data, as in the British charity Full Fact' attempts to debunk false claims in screen captions.

In his 2018 report for the Reuters Institute for the Study of Journalism, Graves essentially details the same two approaches; however, in the case of trained algorithmic models, he puts a special emphasis on *stance detection*, a machine learning technique that tries to figure out whether a claim is supported or not. This technique is mostly studied at a university level at the moment: recently, a research team at the University of Waterloo, in Canada, has built a stance detection tool able to detect fake news 9 times out of 10 (Waterloo News, 2019). That being said, a MIT research team pointed out that stance detection tends to reflect our own biases towards language (Gordon, 2019): for instance, negative statements are considered to be more likely to convey inaccurate views, and affirmative ones are more often associated with truth, thus necessitating new models to be developed.

1.2.2   The rise of automated news

In addition to data mining techniques, Diakopoulos also listed automated content production and newsbots as another way of using algorithms for news production purposes. Although they figure in two different chapters in his book, automated content production and newsbots can essentially be grouped together as they fall under the umbrella of "automated news", a computer process generally understood as the auto-generation of journalistic text through software and algorithms, with no human intervention in-between except for the initial programming (Carlson, 2015; Graefe, 2016). Automated news—which is also sometimes referred to as "automated journalism", "algorithmic journalism" or "robot journalism"—relies on a basic utilisation of Natural Language Generation (i.e., NLG), a computer technique that has been used for several decades to generate text in areas like sports, finances and weather forecasting (Dörr, 2016). In the case of automated news, NLG algorithms are used to fetch information on external or internal datasets, this in order to fill in the blanks left on pre-written text. This resembles a bit the game "Mad Libs" (Diakopoulos, 2019), as programmers or editorial staff need to come up with

templates that, on the one hand, include enough elements that can be predicted in advance and, on the other hand, can be connected to a big enough data flow.

Because of these limitations, only a small range of stories can be automated this way, for instance election results, financial news or sports summaries. Although there is little machine learning involved at the moment, this is becoming a growing area of interest: some machine learning applications of NLG production are already being advertised on the websites of companies that specialise in delivering automated content to business, media, and governmental organisations alike (Narrativa, no date 1); the European Union-funded project EMBEDDIA is looking at including elements of machine learning in automated news generated using pre-written templates to make it less formulaic and nicer to read (Leppänen, 2019; Rämö & Leppänen, 2021); and the Czech news agency ČTK has been experimenting with machine learning techniques to generate automated news templates, with the help of a research team at the University of West Bohemia (Stefanikova, 2019).

Finally, it is worth noting that, now, the use of the term "automated journalism" can refer to automated audio broadcasts (see Heiselberg, Blom & van Dalen, 2022) and automated news videos as well (see Thurman, Stares & Koliska, 2022), although it is important to assess whether these remain NLG-to-audio and NLG-to-video technologies, without any human intervention in-between. There are also text summarisation efforts that are being made as of late, as in the Associated Press' partnership with the firm Agolo to transform journalistic text into automated summaries for broadcast or social media (Marconi & Siegman, 2017), or in the Bavarian public service broadcaster Bayerischer Rundfunk's collaboration with the Technical University of Munich to turn written material into summaries, polls or quizzes for social media (Döllerer, 2021). However, I have decided not to include automated summarisation in my research as it does not feed on any new journalistic information.

Automated news started to be more discussed in the 2010s as *The Los Angeles Times* began covering homicides in an automated fashion (Young & Hermida, 2015) and launched a tool to generate earthquake alerts (Schwencke, 2014), while The Associated Press partnered with the firm Automated Insights to automate corporate earnings stories (Colford, 2014). Proponents of automated news typically develop

their technology in-house, outsource it to an external content provider or use third-party solutions that let journalists design their own automated stories. For instance, the *Washington Post* developed an in-house tool to produce short automated pieces during the 2016 Rio Olympics (WashPost PR Blog, 2016); *Le Monde* collaborated with the firm Syllabs to automatically cover the results of the 2015 regional elections in France (Rédaction du Monde.fr, 2015); and the BBC subscribed to an online platform, Arria NLG Studio, that lets journalists template out their own automated stories using a type of No-code language that makes it accessible to editorial staff with little computing experience (Molumby & Whitwell, 2019).

As for its types of usage, automated news can be used to publish simultaneously at scale, as the Swiss media group Tamedia did with the generation of almost 40,000 hyperlocal stories to report on the outcome of a referendum (Plattner & Orel, 2019), or serve as first drafts to assist journalists with their own writing, as this seems to be the case at *Forbes* and at the *Wall Street Journal* (Willens, 2019; Zeisler & Schmidt, 2021). "Human-in-the-loop" approaches like these are especially relevant when handling edge cases as, according to Broussard (2018, p. 177), "there are things that a human can see that a machine can't". Sometimes, automated news is generated seamlessly as it connects to the organisation's content management system; at other times, it requires setting it up elsewhere and clicking on a command to generate it, which still counts as automation as no human intervention is made to the text. This also applies to automated visualisations that can come with text.

1.3   THESIS OUTLINE

In this introduction, I have outlined how the relationship between technology and journalism goes beyond Pavlik's assumption that "journalism has always been shaped by technology". Instead, as shown by Schudson and Örnebring, this relationship is more complex: mass media also drive the demand for technological innovation and technology can be leveraged to reflect the needs of a value system, which in turn contributes to shape journalism practice. Besides, as reported by Powers, news workers can react to new technological capabilities being brought into the newsroom in three ways: first, by seeing them as an extension of occupational practices; second, by perceiving them as a threat that needs to be subdued; third, by considering them

to be fertile ground for the reinvention of occupational norms and practices. In this PhD dissertation, I will take all these reflections into account to better explore the perceived impacts of algorithmic news production on media practitioners and journalism practice as a whole. As its most developed application, automated journalism is an adequate case study to investigate and will remain the focus of this research thesis, which makes a unique contribution to the field of journalism studies and to social science in several ways.

First, this study is—at the time of writing—the most comprehensive overview of the deployment of automated news across a wide range of organisations and in many countries: as part of this research, contact was made with 33 media practitioners, technologists and executives working at 23 news organisations that are based in 13 countries, thus expanding on earlier exploratory work that used content analysis to have a wider view of how automated journalism is implemented within newsrooms (van Dalen, 2012; Carlson, 2015). To be able to critically investigate the changes it brings to journalism practice, I rely on a distinctive utilisation of Actor-network theory to map out the main transformations that it is going through, combined with Bourdieu's Field theory to be able to tell the power dynamics that this creates. More importantly, my thesis posits that working with automated news reveals the existence of new form of cultural capital that news workers need to acquire, which I call *distinct-abstract* capital in that they need to conjugate the specifics of journalism practice with the kind of abstract reasoning that is at the core of computer programming.

In chapter 2, I will follow up on Örnebring's lead and address the wider societal context that surrounds the adoption of automated news—be it technological (i.e., datafication and renewed interest in artificial intelligence) or journalistic (i.e., tensions between journalism practice and technology adoption). Then, I will look at the main debates in automated journalism research so as to determinate an appropriate research gap.

In chapter 3, I am introducing Actor-network theory and Bourdieu's Field theory as my two frameworks of choice, then detail how I went on to further operationalise Field theory as it is most suited to investigating critical aspects of automated news.

Chapter 4 summarises the methodology used in this PhD dissertation, from ontological positions to a detailed account of the procedures that I have chosen to follow.

In chapter 5, I start off my empirical inquiry with a descriptive account of how automated news was used to cover COVID-19, how the BBC experimented with it in 20219 and how it was employed across a wide range of media organisations based in different countries.

In chapter 6, I proceed to a more critical approach where I investigate the way media practitioners react to automated news, by looking first at the challenges of having to rely on external datasets to set it up, then at the necessity for media practitioners to develop a "computational thinking" mindset and, third, at the tensions occurring between editorial staff and, one the hand, technologists in news and, on the other hand, players external to journalism, in this case Big Tech companies.

Finally, in chapter 7, I answer the research questions I have set out in this dissertation and reflect on their main takeaways as well as on this thesis' contributions to theoretical considerations and journalism studies. I also list down practical recommendations for more industry-focused applications of this research thesis' findings.

2   LITERATURE REVIEW

In the introduction, I have delved into the complex relationship between journalism and technology, demonstrating that, on the one hand, it involves a mutual shaping relationship and that, on the other hand, the way technology is used can be influenced by societal context (i.e., capitalism), which in turn results in shaping journalism practice. In this chapter, I will first reflect on the latter while looking at the wider technological and journalistic context that surrounds the adoption of automated news, then introduce a review of automated journalism research to identify an appropriate research gap and to come up with my research questions for this study.

2.1   DATAFICATION AND NEW "SPRING OF AI"

To all appearances, the societal context that surrounds the adoption of automated news seems to be partly revolving around two technological changes that have been occurring over the last decades and that have provided fertile ground for the development of this technology: first, data collection and processing at scale—also known as *datafication*—second, the use of ever more advanced algorithms to be able to interpret this wealth of data. In this section, I will give a brief overview of both phenomena before bringing in critical considerations that relate to these technical breakthroughs.

2.1.1   Increased datafication in all strata of society

Mayer-Schönberger and Cukier's (2014) definition of datafication constitutes a good starting point to reflect on data collection and processing at scale. Departing from the word "data" in its original sense—which in Latin means "given" as to express a fact that is taken for granted—they stress that the use of this word now rather refers (p. 78) to "a description of something that allows it to be recorded, analyzed, and reorganized". Although Mayer-Schönberger and Cukier ackowledge that collecting

information on citizenry has been going on since biblical times—for instance through censuses or underwriting risk in insurance policies—they nevertheless bring our attention to a new process whereby "all things under the sun" are being collected and rendered in a digital form:

> There's no good term to describe what's taking place now, but one that helps frame the changes is *datafication* (...). It refers to taking information about all things under the sun—including ones we never used to think of as information at all, such as a person's location, the vibrations of an engine, or the stress on a bridge—and transforming it into a data format to make it quantified.
>
> (Mayer-Schönberger & Cukier, 2014, p. 14)

Today, this process of datafication is mostly visible to us in the domain of online advertising, as a wide range of web metrics—like click-through rates and average time spent on page, but also demographic information, location and past online behaviour—are being used to target potential clients and generate profits. This is especially true in the platform economy, as Big Tech companies closely monitor and collect users' data to monetise it, in addition to controlling access to it *via* application programming interfaces or APIs (van Dijck, Poell & de Waal, 2018). This datafication turn has become even more acute in recent years with the adoption of mobile devices that can be paired with location-based services (i.e., GPS, cellular masts, WiFi network) or that are equipped with sensors able to detect movements (i.e., accelerometers, gyroscopes, compasses) as well as surroundings (i.e., barometers, light and proximity sensors). Moreover, the use of these devices has given rise to a form of data collection even more centred around an individual's own behaviours and characteristics: *lifelogging* or the *quantified self*. If *lifelogging* speaks to recording activities using wearables (e.g., a pair of smart glasses) in order to constitute a sort of visual memory of someone's life to pass it on to future generations (Gurrin, Smeaton & Doherty, 2014; Lupton, 2016), the *quantified self* rather refers to a more calculation-heavy approach of "using numbers as a means of monitoring and measuring elements of everyday life and embodiment" (Lupton, 2016, p. 9), for instance keeping a tally of performances when exercising. By the mid-2010s, the *quantified self* movement was considered to be only followed by (*ibid.*, p. 94) "a

disparate group of fitness aficionados, medical maniacs, and tech junkies", but now a single look at all the wrist-worn devices to measure sleep or health available on the market clearly points to a surge of interest in these types of products, and in the "quantified" way of life that comes with it.

Finally, the new frontier of datafication can most certainly be found in the recent push to move towards *linked data*, a new form of Internet governance that aims to develop a common set of standards when creating metadata, which can then be rendered under nodes and relationships in *knowledge graphs* (see Figure 1). The primarily goal of *linked data* is to make content on webpages more machine-readable and better interconnected with other material, so as to have a much more holistic view of a person while linking unconnected pieces of information or open data. As such, *linked data* has been advocated for by Internet pioneer Aaron Swartz (2013, p. 54) who, shortly before his death, touched on the idea of "letting software use the vast collective genius embedded in its published pages". Now, one of its main proponents is the founder of the World Wide Web, Tim Berners-Lee, who sees in *linked data* the opportunity for users to take back control over their own data.
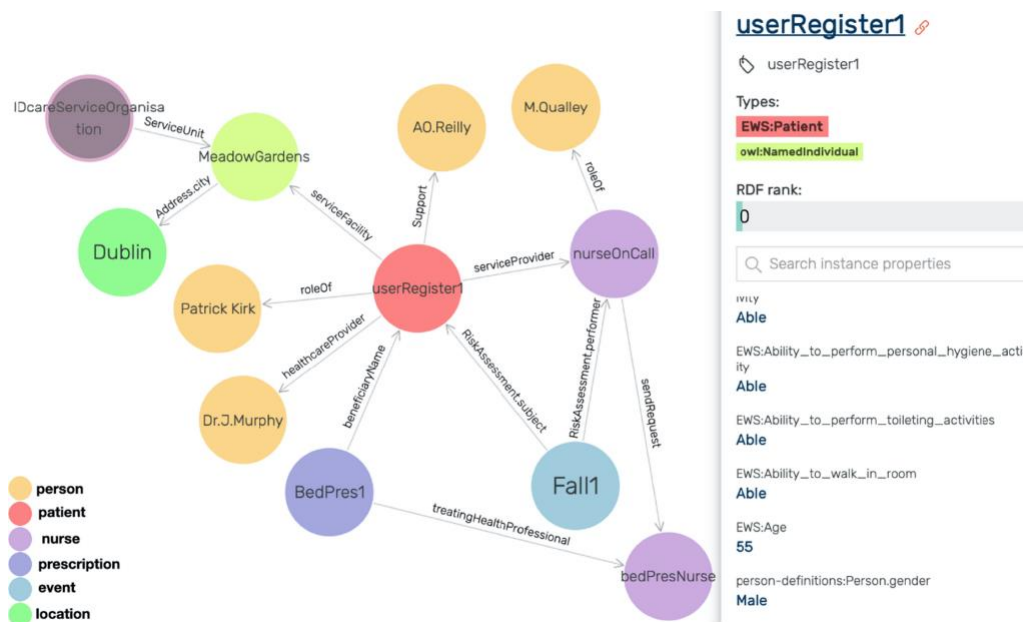
Figure 1: Knowledge graph example. A "knowledge graph" helps visualise linked data through nodes and relationships between entities, as in this mock-up example of a knowledge graph that can be used in nursing. Source: Hussey, Das, Farrell, Ledger & Spencer, 2021.

2.1.2 Towards more advanced machine learning techniques

In parallel with mass data collection and processing, the use of ever more advanced algorithms to be able to interpret this data—for instance while making classifications or, more importantly, predictions based on it—has become even more critical. Over the years, the development of these types of techniques has stirred up a debate around two approaches, which have come to shape the field of artificial intelligence as a whole (Cardon, Cointet & Mazières, 2018): first, the *symbolic* approach or "Good Old-Fashioned Artificial Intelligence" as Haugeland (1985) called it, which is based on specific rules being encoded in expert systems (i.e., computer systems that emulate human decision-making abilities); second, the *connectionist* approach, which stands for computer systems mimicking the functioning of the brain while processing data through several layers of artificial neurons (generally through attributing a weighting score), and which are also known as *neural networks*. To a certain extent, template-based techniques that largely prevail in automated news at the moment (see section 1.2.2) can fall under the umbrella of *symbolic* expert systems.

Although symbolic approaches have been at the forefront of artificial intelligence research from the mid-1950s to the end of the 1990s—and culminated with IBM's computer system Deep Blue beating chess champion Garry Kasparov in 1997—connectionist approaches have since made a comeback, most notably in the domain of computer vision after Krizhevsky, Sutskever and Hinton introduced a whole new benchmark in 2012 with a new type of neural networks that was made possible using graphical processing units (GPUs), a computer chip that was so far employed to power present-day video games[7]. Just as in symbolic approaches, one of the latest landmarks in connectionist AI also has to do with a computer beating a world-renowned champion at his own game: in 2016, a program developed by Google's subsidiary DeepMind—which made use of neural networks and more

---

[7] In 2018, the *New York Times* reported that high demand for GPUs was putting a strain on available supplies, not just because of piqued interest in neural networks, but also because of cryptocurrency mining, which requires the same kind of high computing power that is provided by this type of hardware (Metz, 2018).

symbolic tree search methods (Silver *et al.*, 2016)—managed to beat champion Lee Sedol at the extremely complex game of Go.

After two periods known as "AI Winters" in the 1970s and at the turn of the 1990s (Cardon, Cointet & Mazières, 2018), the field of artificial intelligence is currently undergoing a new "Spring" that is mostly related to growing interest in *machine learning*, a set of computer techniques that essentially rely on three kinds of statistical calculations: one known as *supervised learning*, which employs mathematical formulas such as linear regression and k-nearest neighbours classification to make predictions out of a wide range of data; another one known as *unsupervised learning*, which makes use of other formulas such as clustering and principal component analysis to interpret patterns based, again, on mass data collection; and finally, one known as *reinforcement learning*, an elaborated technique where both predictions and interpretations are put in confrontation in a sort of iterative loop until a desired output is reached.

An example of this last type of learning can be found, for instance, in generative modelling with the development of adversarial neural networks (see Figure 2): in this case, a "generator" algorithm processes training data, such as a database of real people's faces, in order to *predict* what a generic face would look like just by making inferences on some of their traits; a synthetic face generated this way is then passed on to a "detector" algorithm, which makes use of similar neural techniques to detect whether it is a fake or a real person's face, and forwards its feedback to the generator algorithm so that it can compute a more realistic face to be evaluated again. This goes on in a sort of iterative loop until the detector algorithm mistakes a computer-generated face for an real one, in which case the desired output has been reached.
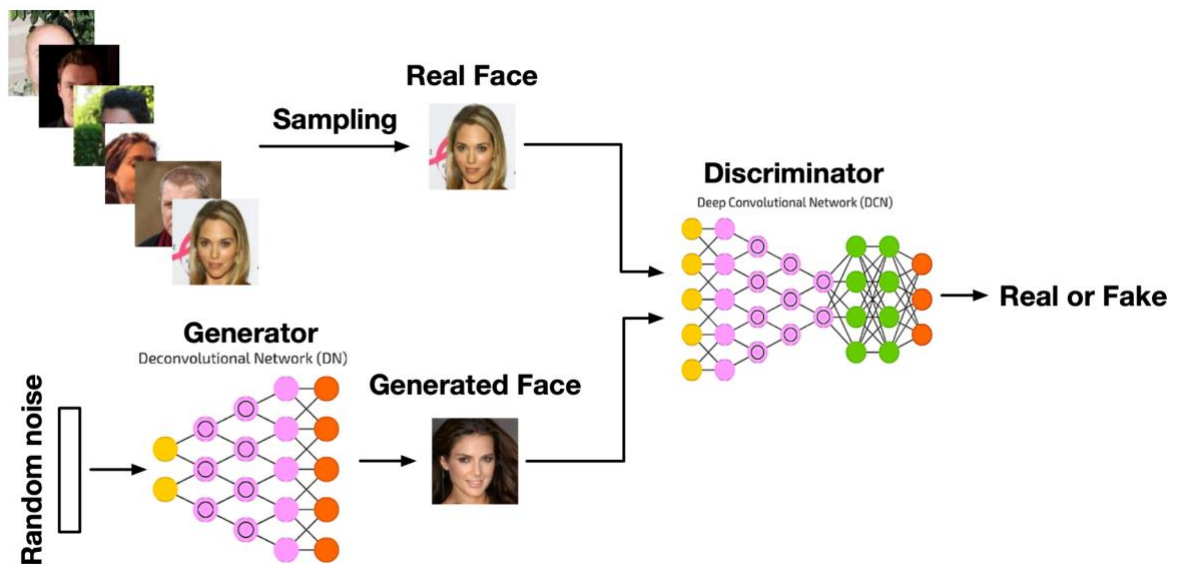
Figure 2: GAN architecture. In a generative adversarial network (GAN), a generator algorithm (bottom) processes training data so as to come up with a synthetic face that is then evaluated by a detector algorithm (right)—along with real people's faces (top). Feedback is sent to the generator algorithm every time a synthetic face is labelled as "fake" by the detector algorithm, so that it can compute a more realistic face to be examined again. Success is achieved when the detector algorithm mistakes a virtual face for a real one. Source: Wang, She & Ward, 2020.

These reinforcement learning techniques have made possible the generation of "deep fake" videos, which, most of the time, are used as spoofs to depict celebrities in a fun way, like representing American actor Nicolas Cage under the traits of an Austrian governess in the 1965 movie *The Sound of Music* (Thorfinn Karlsefni, 2019). That being said, *deepfakes* also bears more concerning impacts in relation to the generation of fake news, like when the environmental group Extinction Rebellion Belgium produced a deep fake video showing Belgian prime minister Sophie Wilmès giving an official speech linking COVID-19 and previous pandemics to human's exploitation and destruction of the environment (Galindo, 2020) or, even more worryingly, when another one depicted Ukrainian president Volodymyr Zelenskyy telling his troops to stop fighting Russia, even though in this case it was of poor quality (Allyn, 2022).

2.1.3 Datafication and AI: critical concerns

If anything, the development of deep fake technologies demonstrates the need to take a look at the less positive sides of all these technical breakthroughs. For instance, Zuboff (2019) argues that collective ignorance of personal data being processed for commercial purposes paves the way to a form of "ubiquitous rendition" and, consequently, to a new form of *surveillance capitalism* that revisits Adam Smith's allegory of the *invisible hand*: according to Smith, *unknowable* market dynamics call for free enterprise and less state regulation so that an *invisible hand* can then arrange a way to the "wealth of the nations". In Zuboff's argument, though, the form of mass data collection and processing that is being done by *surveillance capitalists* gives them a competitive advantage, for they are able to access detailed information that helps anticipate the *invisible hand*'s actions. At the same time, they actively seek out the same kind of unobstructed marketplace that Smith wanted, so that they can continue exploiting what Zuboff (p. 498) calls *surveillance capital*, made-for-profit material that "derives from the dispossession of human experience, operationalized in its unilateral and pervasive programs of rendition":

> The combination of knowledge and freedom works to accelerate asymmetry of power between surveillance capitalists and the societies in which they operate. This cycle will be broken only when we acknowledge as citizens, as societies, and indeed as a civilization that *surveillance capitalists know too much to qualify for freedom.*

(Zuboff, 2019, p. 499)

Another important concept to talk about is *dataveillance*, a term coined more than 30 years ago by information technology specialist Roger Clarke to address the more coercive aspects of scrutinising records, but which has been reinterpreted ever since to include digital practices as a whole. In this sense, Raley illustrates how dataveillance is well concealed when she remarks (2013, p. 125) that "the syncing of browser history with personal and application data has successfully and for the most part uncontroversially been situated under the rubric of 'enhanced user experience'". Besides, van Dijck (2014) observed that dataveillance practices did not fall only on

Tech Giants' shoulders, but also involved a whole apparatus that included government agencies, as this proved to be the case when the Snowden Files revealed that the National Security Agency, in the United States, was collecting social media user's data for surveillance purposes.

Lastly, Couldry and Mejias (2019a) suggested that, to a certain extent, surveillance practices linked to mass data collection and processing can be related to the form of continuous appropriation that characterised historical colonialism. Couldry and Mejias introduced the concept of *data colonialism*, which draws a parallel between data exploitation and colonial empires tapping into another country's resources and manpower. According to them, colonial power is no longer to be understood as foreign settlement only, but also as a form of might that is being held by Western technological companies like Amazon, Apple, Facebook and Google, but also Chinese ones like Baidu, Alibaba, and Tencent:

> More explicitly defined, *data colonialism* is our term for the extension of a global process of extraction that started under colonialism and continued through industrial capitalism, culminating in today's new form: instead of natural resources and labor, what is now being appropriated is human life through its conversion into data. The result degrades life, first by exposing it continuously to monitoring and surveillance (through which data is extracted) and second by thus making human life a direct input to capitalist production. Data colonialism is, in other words, an emerging order for appropriating and extracting social resources for profit through data, practiced via data relations.
>
> (Couldry and Mejias, 2019b, p. XIX)

As for recent machine learning models, concerns about maintaining stereotypes against marginalised communities in the way training data are selected and set constitute yet another source of preoccupation, which is further reinforced by a failure to recruit specialists with a diverse background in Silicon Valley. Such observations are shared by Gebru (2019), who noticed that not only these types of algorithmic biases are not addressed because they do not affect the dominant group in the industry, but also that they contribute to further targeting oppressed groups. She takes an example from machine translation: a Palestinian was detained for a few hours by Israeli Police after his Facebook post that said "good morning" in Arabic was

translated into "attack them" in Hebrew and "hurt them" in English. Gebru (2019, p. 22) argues that "structural issues at play" made this embroilment possible, as "the oppression of Palestinians also makes it more likely that whatever translation errors that do exist are more harmful towards them". She goes on to say (*ibid.*, p. 20) that "tools used by Google and Facebook currently work best for translations between English and other western languages such as French, reflecting which cultures are most represented within the machine learning and natural language processing communities". To her, this becomes even more of a problem with projects like the Extreme Vetting Initiative that was once brought forward by the United States' Immigration and Customs Enforcement agency, which envisaged at a time using machine learning to make predictions on whether foreign visitors could positively contribute to the country or were, instead, at risk of committing criminal or terrorist acts (see Harwell and Miroff, 2018). She deplored that only a few groups within the artificial intelligence community were truly speaking up against proposals like this, and that the (p. 24) "extreme underrepresentation of marginalized groups in the latter community makes it even more difficult for them to care".

In this section, I gave a brief overview of datafication and recent advances in machine learning, which I am connecting to a set of critical considerations to show their downsides as well. As automated news can be understood as a technological development that is directly grounded in this socio-technical context (e.g., growing access to open data and machine learning techniques to improve automated news), it is therefore essential to investigate it with a critical eye. However, the societal context that surrounds the adoption of automated news would not be complete without having a more thorough look at the way media practitioners reacted to new technological capabilities being brought into the newsroom, which I am going to do next.

2.2   INVESTIGATING THE JOURNALISTIC CONTEXT

Now that I have established the wider technological context that surrounds the adoption of automated news, I will look at the more journalistic context that saw technology being introduced into the newsroom. This is no easy feat as it could be

traced back to the invention of print technologies and as it may involve, as seen in the Introduction, a certain degree of technological determinism. To mitigate this, I will rather show that, instead of technology, political influences dominated the age of print and television news, before a shift occurred towards more audience and advertising types of pressures in the digital era.

2.2.1   Print and television: political and commercial pressures

Studies on newsroom practices have been envisaged as early as the beginning of the twentieth century, when Weber presented at the first Congress of Sociologists in 1910, in Frankfurt, his vision for a "sociology of the press". He proposed to examine not only "the product at hand" (i.e., newspapers), but also "the fate and the situation of journalism as a profession" (Weber, 1976, p. 100). To conduct this massive study that would have potentially encompassed the German, French, English and American media markets, Weber considered using questionnaires to collect information on the journalists themselves and on newsroom practices of the time (Weber, 1998). However, Weber's plan was thwarted after he got embroiled in a legal dispute with a newspaper, making him think that journalists' distrust would be too much of an impediment to properly conduct the research (Hennis, 1998). Consequently, empirical work into newsroom practices only started in the 1950s, and on a much smaller scale than the one envisioned by Weber.

Two fundamental studies can be seen as laying the groundwork for this type of investigation: one conducted by Manning White in 1950 and another carried out by Breed in 1955 (see Reese and Ballinger, 2001). While Manning White analysed the thought process that guided a wire editor' selection and rejection of news items in a local newspaper, Breed rather focused on hidden mechanisms of control that help publishers ensure that journalistic staff abide by their editorial policies. The influence of the political sphere onto journalism practice thus begins to be noticeable.

In Manning White's article—the first of its kind in gatekeeping media studies—this form of power wielded by the political sphere is visible in strong political statements made by "Mr. Gate", the observed wire editor. Coming back to

the reasons as to why he discarded a wire story over another[8], Mr. Gate indicated in a number of cases that he considered this to be "propaganda". In another instance, he shared that a person mentioned in a story was being "too Red". In addition, Manning White also observed (1950, p. 388-389) that "in almost every case where he had some choice between competing press association stories Mr. Gates preferred the 'conservative[9]'". The wire editor's political orientations were verified at a research stage where he answered follow-up questions:

> I have few prejudices, built-in or otherwise, and there is little I can do about them. I dislike Truman's economics, daylight saving time and warm beer, but I go ahead using stories on them and other matters if I feel there is nothing more important to give space to. I am also prejudiced against a publicity-seeking minority with headquarters in Rome, and I don't help them a lot.

> (Manning White, 1950, p. 390)

Beyond his own subjective views, Mr. Gates also recognised prioritising stories "slanted to conform to our editorial policies" (*ibid.*, p. 390), thus revealing the influence of the political field that is made through the publisher's editorial line. It is this aspect that Breed set out to explore in his article on the forms of social control that make journalistic staff abide by these policies. Breed essentially lists out six forces that keep the "potentially intransigent staffer" at bay (1955, p. 329). These are: *institutional authority and sanctions*; *feelings of obligation and esteem for superiors*; *mobility aspirations*; *absence of conflicting group allegiance* (e.g., absence of a union combatting this policy); *the pleasant nature of the activity* (i.e., editorial staff stay in the newspaper business when they could go after better paid opportunities); and, finally, *news become a value*, meaning that a go-getter attitude among journalists to get news as fresh as possible takes precedence over challenging their publisher's line. Although there is room for manoeuvre amidst those (e.g., taking advantage of foggy areas within the publisher's own policy, or even leaking stories to a competitor to see

---

[8] The wire editor was asked to comment on each of the copies he rejected every day, over a seven-day period.

[9] Manning White speaks of "conservative" political leanings, but also of Mr. Gates' preference for a more conservative writing tone.

them released elsewhere and then make the case for publication), they nevertheless testify of the influence of the political sphere on journalism practice.

In addition to these early empirical investigations into print newsrooms, it is also worth looking at another stream of research that documented journalists' work routines in the 1970s and 1980s, while focusing by and large on the impacts of television. Similarly to Manning White and Breed, these scholarship reasserted the prevalence of political influences on journalism practice, although economic influences driven by audience and advertising considerations started showing. To begin with, newsroom researchers like Gans (1979), Tuchman (1978) and Fishman (1980) have illustrated how—for pragmatic, practical and efficiency reasons—media practitioners heavily relied on established sources, thus further reinforcing their legitimacy.

For instance, Gans (1979, p. 282) describes the process through which journalists choose powerful sources, assuming that they are "the most easily and quickly available, as well as most reliable and productive", and also "the most efficient"[10]. Gans also remarked (*ibid.*, p. 283) that "it is efficient for journalists to respect the power of sources", because "if they did not regularly choose influential sources, news firms and journalists would have to spend time and monies to fend off their pressure". Likewise, Tuchman (1978, p. 212) detailed a similar way of working whereby "news processing is itself routinized according to the way occurrences at legitimated institutions are thought to unfold". According to her typology of story formats (Tuchman, 1973)—where "spot news" covers unanticipated events (e.g., breaking news) and "continuing stories" those that unfold over time (e.g., parliamentary proceedings, lawsuits, etc.)—she observed (1978, p. 212) that, for efficiency reasons, "predicting the course of continuing stories at legitimated institutions enables editors to plan which reporters will be available for spot-news coverage on any one day". Fishman (1980) goes even further to say that news organisations benefit from what he calls a "welcomed subsidy", which takes the form

---

[10] Here, efficiency must be understood not in terms of pure economic benefits, but as the most rational way to allocate staff, air time, print space or production time to deliver news as fresh as possible, at a given pre-set time.

of informational material prepared in advance by institutions and that comes at no cost for them:

> Imagine the labor costs to a news organization if it did not have such bureaucracies to rely on for this essential work. The organization simply could not cover the scope of news it is accustomed to presenting without a massive increase in labor costs. In effect, an enormous network of governmental agencies, corporate bureaucracies, and community organizations underwrite the costs of news production. The modern news organization is predicated on this invisible subsidy.
> Under inflexible deadlines and expectable increases in story quotas, reporters and editors cannot resist the preformed, pre-scheduled, and factually safe raw materials that bureaucracies provide. By propping up advertising revenues and at the same time holding down labor costs, news organizations create the work conditions which necessitate a reliance on the free services of agencies outside the newsroom.
>
> (Fishman, 1980, p. 151-152)

In the end, these processes of facilitating journalistic work can be seen as further legitimising existing power structures. As an illustration of this, Fishman argues (*ibid.*, p. 152) that, in exchange for these "free services", the media "bestow on routine news sources equally valuable services: publicity and legitimation". Tuchman (1978, p. 212) witnesses the same dynamics when journalists seek out reactions to a news story since they generally go after the same legitimated sources like "governors, mayors, presidential aspirants, senators, other legislators, and quasi-legitimated leaders". To her, chasing Republicans and Democrats for comments not only embodies the current political landscape, but also participates in further legitimising it.

Along with political influences, other forces driven, this time, by economic considerations like audience and advertising pressures also penetrated the area of journalism. Even though Gans (1979) touched on "audience power", he held that it had less influence than "source power", as according to him the news media were providing audiences with a great deal of stories they did not care about, thus making him think that they rather acted as "spectators" who were unable to shape news content. He went on to suggest (*ibid.*, p. 284) that, in theory, audiences should have

more influence than power sources—for they could choose to reject all of these stories all at once—but practically speaking "the audience could exert such power only by a mass boycott, requiring the nearly impossible task of transforming itself from a heterogeneous aggregate of spectators into an organized and mobilized movement".

By contrast, Epstein (1973) viewed audiences as having a greater sense of agency. According to him, national television news does not have a significant audience of its own and mostly inherits viewers from preceding programs. This, in turn, means that should producers engage more overhead costs to widen the scope and quality of news, this may not necessarily translate into more viewers and advertising revenues. As Epstein put it (*ibid.*, p. 260), "there is no economic incentive to spend money on searches for original information, or intelligence gathering, since it is not presumed that scoops, exclusives or original reporting significantly increase the audience, and hence the revenue, for network news". Consequently, he notes, these networks rather rely on stories already published in the *New York Times* or that are dispatched to them through wire services or *via* other sources. This situation echoes the lack of incentives to invest in investigative reporting, both in print and television news (see Hamilton, 2018), as this format generally comes with a lengthy amount of fieldwork and research.

On a somewhat discursive note, it is worth mentioning that, in his study of televised news, Altheide (1976) demonstrated that the news making process acts as a medium in itself, a medium through which reality is filtered out and distorted so as to be able to hide all of the influences—including commercial and political ones—that shape media practitioners' views on the world. Altheide introduces the concept of "news perspective", which stands for the idea that, for practical purposes, events are being framed as a narrative with a beginning, middle and end. As he writes (*ibid.*, p. 173-174), "rules of thumb, editing techniques, marketing research, use of themes and angles, and even writing a story to fit with another to make the 'show flow,' has become the rational way to change the world in order to present it as news". In a way, this bears resemblance to the form of technological determinism portrayed by McLuhan (see section 1.1), where the medium itself carries a power of its own—although in this case it is to better hide away all of the external influences that shape journalists' minds. Altheide further observed that the "news perspective" sits in a

position of power, since it has become standardised practice in the industry and translates into job security, praise among peers and being awarded journalistic prizes.

2.2.2   Tensions around metrics in digital media

Looking now at the digital shift in the media industry that happened over the last few decades, two major forms of external influences appear to be visible: on the one hand, the aforementioned form of influence that is linked to audience and advertising pressures and, on the other hand, another external force driven, this time, by the growing importance taken by technologists and technological companies within newsrooms. As a matter of fact, the audience power described by Epstein (i.e., television ratings) seems to have evolved into the use of digital metrics to calculate, for instance, the number of page views or the amount of time spent on a webpage (see section 2.1.1). To comprehend these transformations, though, it is first necessary to have an good understanding of how the Internet reshaped the relationship between journalists and their audiences.

In her ethnographic study of the *New York Times* in the digital era, Usher (2014) documented how pressures stemming from immediacy—a core ideal of the World Wide Web—translated into difficulties to conciliate the newspaper's print heritage with the perceived need to produce "24/7 ASAP news". As she noticed (*ibid.*, p. 232), "journalists did not know what to make of what seemed like an imperative to be responsive to the networked society". Usher observed that journalists adopted new work routines, like writing contextual information ahead of time to get the news out as soon as possible and then updating it throughout the day. She also detailed that, although home page editors seemed to have a sense of how to manage their workload, they were rarely seen leaving their desks because of a non-stop flow of potential news stories. Usher warned against the risks associated with what she sees as "churnalism," "hamsterization" and a "news cyclone", conditions under which journalists are pushed (*ibid.*, p. 233) to "produce content right away, rather than taking time to think carefully about news judgement". This was visible, for instance, in a situation where journalists rushed into interpreting unemployment figures: they understood first that these trends were bad for the United States' economy, when the outlook was actually positive.

Amidst these changes, information and communication technologies also enabled a direct line of communication between journalists and their audiences, sometimes resulting in their work being challenged more easily. In his ethnographic study of the deployment of digital news at three newspapers (including the technology section of the *New York Times*), Boczkowski (2004) observed that online access made it possible for readers to have a say in shaping news content: they could transmit their views to journalists by email, while writing comments on the newspaper's online forum or when publishing their own newsletter on its website. Although it stands true to say that this is a more conversational approach to news making, Boczkowski also reckoned (*ibid*., p. 186) that it "opens the news to a higher degree of contestation, expressed either by direct conflict of opinions or indirect multiplicity of views, than the typical case of traditional media." That being said, in their analysis of user-generated content on 12 national United Kingdom newspapers' sites, Hermida and Thurman (2008) observed that traditional gatekeeping still remains strongly in place, since the vast majority of these websites prevent readers from posting comments directly on it without any prior check. According to them, this reflects the limited risk-taking nature of newspapers as well as editors' concerns for trust, reputation and legal issues (see Thurman, 2008).

Coming back to the use of digital metrics in recent years, this form of audience pervasiveness is most visible in the involvement of web analytics companies providing newsrooms with key indicators on performances (see Belair-Gagnon & Holton, 2018). In her ethnographic study of two online newsrooms; one in New York, one in Paris; Christin (2020) remarked that these audience metrics were indeed very much present in the mind of web journalists: they used these numbers so as to get a sense of who their readers are, as opposed to print journalists who rather used market surveys, anecdotal evidences like letters to the editor or their own imagination for that (see Nelson, 2021 for the way journalists perceive their audiences, which remains according to him rather "imagined" than "truly known").

As Christin detailed, these forms of representations gave rise, in both newsrooms, to tensions originating from two modes of evaluation: one being editorial-based, which emphasises quality content and peer approval, and another one being rather click-based, which gives preference to quantitative measures of

success like audience metrics. In her investigation, she uncovered a clash between these two modes of evaluation, as quality articles did not necessarily translate into increased metric scores. As she writes, there could be a connection to be made between these forms of audience pressures and precarious work conditions in journalism, which ultimately results in a diminution of the type of specialised coverage that is generally praised among peers:

> As we saw, when news organizations entered the chase for traffic, they began to publish more and faster. They relied on aggregation and clickbait headlines in order to increase their advertising revenues. They hired large numbers of flexible workers to keep their homepages "fresh." They engaged in relentless efforts to make their content more visible on Google, Facebook, and Twitter. These developments had editorial consequences—including an exponential increase in cat videos, scandal coverage, and contentious opinion pieces.
>
> (Christin, 2020, p. 153)

Besides, Christin observed that differences between French and American journalistic work cultures had an impact on journalists' self-perception *vis-à-vis* web metrics: whereas in the United States strong professionalisation that came as a reaction to early market pressures led to a certain distance between journalists and their audiences, in France—since the Dreyfus affair—journalists rather see themselves as intellectuals whose role it is to contribute to the public debate. As a result, American journalists seemed to be clearer about the fact that web metrics represent market pressures—and could choose to reject them as an encroachment to journalistic autonomy[11]—while French staff writers valued those as an indicator of their own public relevance, even if aware of their market role too. In the latter case, this led to increased pressures as (*ibid.*, p. 155) "staff writers experienced greater stress from the metrics that managers liked to critique."

This state of self-alienation because of metrics was also documented by Petre (2021), who concluded after her fieldwork at the *New York Times* and Gawker that real-time metrics came along with a work intensification that brought stress and

---

[11] By contrast, Tandoc and Foo (2018, p. 53) reported—in their study of BuzzFeed News—that paying close attention to audiences could also be celebrated, as this may offer "an ideal for the field of journalism that is, in general, struggling to keep its audience".

exhaustion among news staff. However, according to her, this work intensification is also driven by journalists' own urge to engage with analytics tools like Chartbeat, rather than coming from managerial pressures only. As Petre put it (*ibid.*, p. 190), "once metrics become deeply institutionalized in journalism, they are capable of intensifying journalistic work, extracting increased productivity, fostering competitive dynamics between colleagues, and producing feelings of alienation". This is connected to what she calls the "traffic game", a non-stop competition where journalists contend with each other—and against themselves—so as to be able to boost their number of article views and social media engagements[12].

These gamified logics could be even more concrete, as shown in Ferrer-Conill's (2017) analysis of the system put in place at the sports news website Bleacher Report, which encouraged writers to achieve high metrics scores while rewarding them with virtual points and badges. Even though the journalists interviewed said they viewed this gamified system in a positive light, Ferrer-Conill wrote that, nevertheless, (*ibid.*, p. 717) "the fact that system dynamics shape their production patterns is a cause for concern". According to him, this resonated with previous controversies surrounding possible exploitative practices at Bleacher Report, like when the news website was, for a time, resorting to unpaid contributors. As Turner Broadcasting acquired the news website in 2012, the open contribution model was slowly phased out and the gamified system went through important changes, but remained visible in some areas. For Ferrer-Conill, this illustrates (*ibid.*) "the problems of mixing data logics with traditional journalistic norms and values", especially as web metrics need to be intertwined with professional standards so as to set a news organisation apart and make it legitimate.

2.2.3 Technologists' growing influence in news making

The second form of external influences that are linked to digital transformation within newsrooms involves technologists, who progressively gained in influence in the relationship they entertain with editorial and business colleagues (see Lewis &

---

[12] Petre also illustrated how metrics could be leveraged to advance journalists' own goals, sometimes even going against management as these metrics can make them aware of their true worth. For instance, she gives the example of Gawker writers invoking metrics to ask for promotions and raises.

Westlund, 2015b for a discussion of the interplay between technologists, editorial and businesspeople in cross-media news work). This relationship was characterised as being "full of tensions" in Nielsen's study (2012), which investigated blogging developments at two Danish newspapers. He suggested that these tensions could accelerate innovation within newsrooms, but also be an impediment to change: Nielsen found that journalists were generally seen as being more sceptical and managers tight with money, while criticisms addressed to technologists rather revolved around a lack of commercial and editorial goals, as they were perceived as being eager to try everything out without necessarily having any rationale in mind. This drove him to say (*ibid.*, p. 973) that, in present-day media, "tensions like these will be integral parts of the internal process as journalists, managers, and technologists work together (and sometimes argue and fight over) how tools like blogs, social networking sites, various applications, and mobile media should be integrated into their organizational practices".

These findings echo Westlund's observations (2011) on tensions between "creatives, suits and techies" when it came to producing mobile media content at the regional newspaper *Göteborgs-Posten*, in Sweden. Even though these tensions were somewhat relatively mild—one of them involving for instance the creation of an Iphone application—Westlund still remarked (*ibid.*, p. 327) that "the techies were battling their historically weaker role in the organisation, associated with performing functions such as fixing bugs", or in other words being considered "a kind of service desk". That said, this was about to change as the newspaper's board wanted to reinforce technologists' status and authority, reallocating some of them so that they sit closer to the editorial department.

Just as technologists are gaining in influence in the newsroom, a technological form of pervasiveness is also perceptible in the way media practitioners upgrade their computer skills, one example being the expert use of social media to gather and verify content online, but also to engage audiences *via* dedicated roles like social media editor or community manager. Looking at the use of social media for crisis reporting at the BBC, Belair-Gagnon (2015) described the increased responsibilities taken on by "tech-savvy journalists" who intertwined social media practices with journalistic work, most notably with respect to the User-Generated Content Hub's strategy, a

team dedicated to verifying user-generated content and dispatching it all across the BBC. As she noted, these new responsibilities also involved teaching colleagues on how to handle social media and participating in editorial meetings at the multimedia desk. According to Belair-Gagnon, the rise of these "tech-savvy" journalists can be explained by societal context, as the broadcaster's efforts to rely on user-generated sources reflects a more general trend to be closer to audiences.

Despite the widespread adoption of social media strategies within newsrooms, these sources seem to be not benefiting from the same level of trust than those in traditional reporting. In their analysis of Twitter usage among Irish journalists, Heravi and Harrower (2016) indicated that, although journalists heavily rely on known and authoritative Twitter sources as part of their work routines, these sources are not necessarily trusted, thus leading them to use more well-established sources outside of social media, like officials or "'real-world' networks". Another type of distrust may also be at play at a more organisational level, as Pignard-Cheynel and Amigo (2019) pointed out that positions like social media editor were faced with quite a high turnover rate, their proximity to marketing and participatory logics making them look more distant to the true spirit of journalism.

Beyond the advanced use of social media, "tech-savvy" journalists can also be engaging with computer programming as part of their daily tasks, thus becoming "computational journalists". Computational journalism can be understood as a discipline that initially addressed the use of advanced software in order to assist journalists with their work routines (see Cohen *et al.*, 2011; Flew *et al.*, 2012; Hamilton & Turner, 2009), but then expanded to cover journalists' abilities to solve problems through abstraction and computing skills (see Diakopoulos, 2011; Gynnild, 2014; Stavelin, 2013), a form of mindset that is also known as "computational thinking" (Wing 2008) and can generally be found in computer science. Computational journalists can therefore include creators of news applications—which feature for instance maps, timelines and charts—where journalistic logic is involved, but also data journalists using their programming skills as part of data scraping activities (see

Karlsen & Stavelin, 2014)[13]. Following Diakopoulos (2019), these computational journalists can be trained in three ways: first, by embedding computational thinkers into editorial environments, as in the Knight-Mozilla Open News program that placed open source development fellows in newsrooms (see Lewis & Usher, 2013) so that they collaborate on developing solutions; second, by training journalists in advanced statistics (see Heravi, 2019), thus following Meyer's (1973) recommendation for computer-assisted reporting (see section 1.1.2); third, by rethinking journalism curricula to have more computational and data journalism degrees available, but also doctoral degrees in computational journalism so as to be able to train educators as well.

Among this group of "computationally minded journalists" (Usher, 2018)—which, again, includes data journalists using programming skills—there seems to be either a desire to break away from prevailing journalistic norms or, on the contrary, a willingness to maintain them. Regarding challenging norms, some computational journalists have been pushing for more open source transparency and participation-oriented reporting, which is quite a split from standard journalism practice where "exclusives" are highly sought-after and where journalists generally act as gatekeepers in news selection. For instance, Parasie and Dagiral (2013) found in their study of the Chicago area that programmer-journalists indeed value public access to information, including sharing code. They would vouch for giving audiences large and easy access to data, and for enabling them to combine it with other elements. As Parasie and Dagiral sustain, these stances are linked with ideals such as believing that technological artefacts represent a leap forward for democracy, which is characteristic of the open software and open government scene in Chicago, often described as liberal and progressive.

Likewise, in his analysis of British mainstream media, Borges-Rey (2016) observed that crowdsourced activities and collaborations outside the newsroom figure among the norms and conventions that data journalists wish to institutionalise in order to legitimate their practice. To him, this also acts as a buffer against

---

[13] Borges-Rey (2016, p. 838) makes a difference between a "daily, quick turnaround, generally visualised, brief form of data journalism" for lighter and entertaining formats and a "thoroughly researched, investigative form of data journalism". This distinction is similar to the "ordinary" and "thorough" styles of data journalism identified in De Maeyer *et al.* (2015).

criticisms—for instance when readers attack their data sources or news angles—as data journalists "embrace this kind of public engagement as a natural part of their news reporting" (*ibid*., p. 842). That said, data journalists may retain control over what they choose to share—or not—coming back in that sense to traditional notions of gatekeeping. For instance, Porlezza and Splendore (2019) revealed that almost all of the Italian data journalists they interviewed had a journalistic background before they acquired programming skills, and understood openness as giving readers a glimpse into journalistic know-how, but not as participating in the process.

Finally, computational journalists may also actually embrace journalistic norms as they stand, this so by emphasising their reporting skills over their programming aptitudes. This is essentially what Karlsen and Stavelin (2014) revealed while observing, for instance, journalists creating news applications in Norwegian newsrooms. They stressed that—even though computational journalists need to be able to master (*ibid*., p. 43) "both the inverted pyramid structure of journalistic storytelling and basic iteration statements found in any programming language"— they still see journalistic knowledge as being more important that programming skills, as those are viewed as a more of a technical tool bag as opposed to the "unbroken tradition of journalism." In a parallel with photojournalism, Karlsen and Stavelin asserted (*ibid*., p. 43) that "as you do not want to hire any photographer as photojournalist—you do not want to hire any programmer to do computational journalism". Besides, the computational journalists they interviewed said they saw using computers just like other journalists use notebooks, telephones or microphones: "Sometimes they write a computer program in C#, sometimes they find, install and use a new software tool to get a job done, and sometimes they use the more advanced features in Excel" (*ibid*., p. 44). In Karlsen and Stavelin, it is therefore understood that computational journalists need to think of themselves as journalists first and foremost:

> To become a programming journalist you must accept that journalism goes before programming. You need to be a journalist "by conviction" to avoid conflict and to be able to thrive in the newsroom. It is important to distance yourselves from the technologists working in the ICT (Information and Communications Technology) department. Your fellow journalists should not be in doubt whether you belong to the newsroom or ICT. You need to bypass

ICT by choosing lightweight technical approaches and find solutions that do not require direct assistance from ICT.

(Karlsen & Stavelin, 2014, p. 44)

In this section, I demonstrated how journalism practice underwent a shift, from being influenced mostly by the political sphere in print and television news to being pressured by audience and advertising types of influences in the digital age, and also by a new form of technological pervasiveness. While a new class of computational journalists seems to be challenging prevailing journalistic norms as they push for more openness and audience participation, the opposite is also true as they can actually embrace these norms, this by giving more importance to journalistic ideals than to computer skills only. As one may expect computational journalists to take a leading role in the development of automated news products due to their computing skills, this is even more so an important aspect to take into consideration. After having studied the wider technological and journalistic context that surrounds the adoption of automated news, I will take a look at the current state of research into this area so as to identify a proper research gap and to narrow in on my research questions.

2.3   RESEARCH ON AUTOMATED NEWS

To be able to have such an outlook, I conducted a systematic literature review that was published on Open Research Europe (Danzon-Chambaud, 2021a), whose findings are reproduced as-is or in a slightly modified form in the section that follows. Conducting such a review contributed to filling an important research gap as— although Graefe and Bohlken (2020) conducted a meta-analysis that focused on readers' perceptions—none had, so far, provided a more complete picture that included scholarship on practice as well. In this review, I concentrated on the key features of a selection of academic articles written on automated news, so as to have a comprehensive overview of the field. To do this, I retrieved documents in a systematic manner, using the following combined search query: "automated journalism" OR "algorithmic journalism" OR "robot journalism" OR "machine-written

journalism" OR "computational journalism"[14]. I searched five databases (i.e., Taylor and Francis, Sage, ScienceDirect, SpringerLink, and Scopus) and looked for material published between 2005 and mid-2020 so as to account for the last 15 years of research, a range that should accommodate this research as it goes back five years prior to the launch of *The Los Angeles Times*' pioneering project on homicide coverage (see section 1.2.2).

Out of close to 500 results, I selected articles written in English and that are based on empirical evidence so as to assess the latest findings in the field, and which presented an exclusive focus on automated journalism. As pointed out in section 1.2.2, I understood it 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 stretched to any type of auto generated text so as to account for recent developments as well (see Lindén, 2017a). In the end, my corpus was constituted of 33 scholarly articles on automated journalism (see Appendix A), a number that I am satisfied with given that automated news was introduced into newsrooms and attracted scholars' attention only quite recently.

To conduct an efficient review of the field, I looked, first, at variables that I could quantitatively measure—such as semantical, chronological and geographical features—then at the research methods, fields of inquiry and theoretical backgrounds at play. To do this, I retrieved every keyword mentioned in the corpus, the years the articles were published online and also the countries they originated from, their domains of investigation and the methods used. Lastly, I collected information on theoretical considerations and on bibliographic references cited more than five times throughout the entire corpus. In a second step, I engaged with and critically assessed all of these meta-data so as to get a more qualitative understanding of the main debates dominating the field. No risk of bias was found, apart from limiting myself to articles written in English, a point that I address in the review.

In addition to this systematic approach, I also examined work did not figure in my search because it did not correspond to my criteria. These cover for instance

---

[14] "Automated news" did not appear to be a term that was very much in use at the time of my search. However, the scholarship retrieved *via* the other keywords seemed to be comprehensive enough to be able to fill that gap.

books, conference papers, industry reports and other scholarly articles that were published in the second half of 2020. They are featured here in the section that immediately follows my systematic literature review.

2.3.1 False "robot" depiction and geographical distribution

First, to investigate the various semantics used in the field of automated journalism, I analysed the different keywords that were used in my corpus (see Figure 3). I found that the most frequent ones referred to "robot journalism" and "automated journalism", two terms that are regularly employed in mainstream media and academia to talk about the computer-generation of news text, but which face criticisms as they do not convey quite an exact meaning.
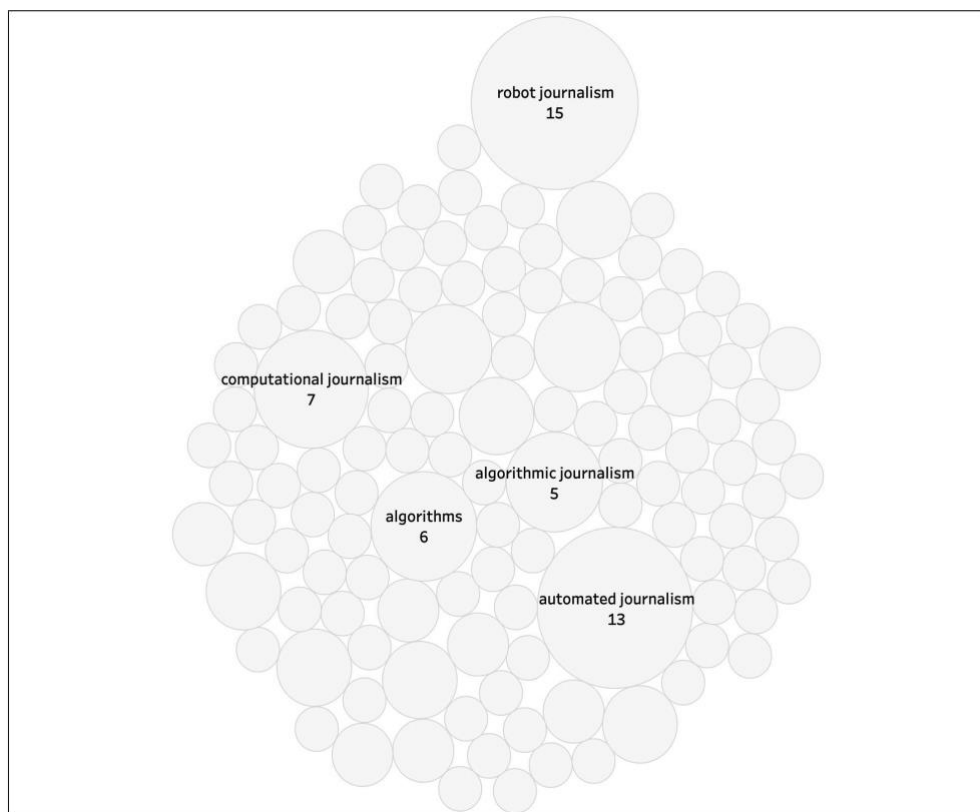
Figure 3: Automated journalism's main keywords. Most-frequent keywords used in corpus. Only keywords mentioned 5 times or more are indicated. Source: Danzon-Chambaud, 2021a.

The most problematic use of "robot journalism" revolves around the fact that NLG involves a computer script and no actual robot (Dörr, 2016). Lindén (2017a, p. 125) actually cautions against a "popular but banal conceptualisation where illustrators more often portray robots writing on computer keyboards", which ultimately plays on journalists' fears of being made redundant and prevents newsrooms from being more innovative (see also Lindén & Dierickx, 2019). This metaphor could even be detrimental to the acceptance of automated news, since readers may ultimately feel deceived after being drawn to believe that this form of technology exhibits some sort of humanness (Waddell, 2018). In addition, there could be no added benefit to using this metaphor since readers have proven to be equally receptive to automated news when it is labelled as an algorithmic product or when it is identified with software names (Waddell, 2019a). Only readers that were previously exposed to robots in fiction pieces seemed to be perceiving it in a more positive light (Waddell, 2018).

That being said, the use of the term "automated journalism" could also be up for debate. Although it is becoming a more preferred term than "robot journalism" in industry and academia, its focus on computer-generated text can be seen as too narrow and does not necessarily reflect media practitioners' own views. In fact, Wu, Tandoc and Salmon (2019a, p. 1453) advance that "automated journalism" can also be inclusive of a whole range of tasks, which involves "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, it is not quite the same for "automated journalism". This should nevertheless be reflected upon, especially as "automated journalism" could also apply computational journalism practices like programming solutions to solve some of the journalists' issues (see Karlsen & Stavelin, 2014). It is also unclear whether it can be employed to talk about automated content other than text (i.e., automated audio summaries, automated video news), with the same caveat that relates to no human participation in-between (see section 1.2.2).

The second type of variables I looked for were the years the articles were published online and the countries they originated from (see Table 1). While no publication was found prior to 2012, a steady growth in the number of articles could be noticed from 2014 onwards, with the only exception of a small decrease in 2015. This picture is

however somehow incomplete as I finalised my data collection in mid-June 2020. This growth can be interpreted as a reflection of automated news adoption, as major outlets made announcements in that sense at about the same time. Looking at the countries from which research originated, the results were not that surprising as I limited my search to English-written scholarship: English-speaking countries (or partly English-speaking in the case of Canada) constituted the largest group (i.e., United States, Australia, Canada, United Kingdom), followed by other Western countries (i.e., Germany, Denmark, Finland, Israel, Netherlands, Spain, Sweden, Switzerland) and a few Asian ones (i.e., South Korea, India, Singapore). No scholarship associated with an African or South American country was found, but this probably had to do with my focus on scholarship in English rather than with technology penetration in some of these areas, although this is not a negligible aspect.

Having made this observation, it is once I combined the online publication dates with countries of origin that I started seeing more interesting patterns. First and foremost, two out of four articles published between 2012 and 2014 originated from Northern Europe (i.e., Denmark and Sweden) while the other half came from Canada and the United States. These two Northern Europe articles can be considered quite pioneering work on automated news, since they first tackled its perceived impacts on the work of media practitioners (van Dalen, 2012) and the perceptions its triggers among audiences (Clerwall, 2014). These findings also suggest that Nordic news outlets had a key role to play in the development of automated news, although much of the spotlight was on—and still remains with—large organisations in the United States like the *Los Angeles Times*, the Associated Press and the *Washington Post*. At the Danish news agency Ritzau and the Swedish media group MittMedia, for instance, automated journalism was introduced as early as 2015. Since they either partially or entirely own the solution they use (see Falk Eriksen, 2018; Lindén & Tuulonen, 2019), Ritzau and MittMedia differ in that sense from other organisations like *Le Monde* and the Associated Press, which adopted automated news at about the same time but outsourced its development to an external NLG provider.

Table 1. Articles' online publication years and countries of origin. Source: Danzon-Chambaud, 2021a.

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

The combination of 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, 2019), others originated from India (Visvam Devadoss, Thirulokachander & Visvam Devadoss, 2019), Singapore (Tandoc, Lim & Wu, 2020) and partly from China (Zheng *et al.*, 2018). We can only assume that this represents the "tip of the iceberg" as it is likely that additional research has been published in local languages that I could not read. In the case of China, for instance, only a handful of information is readily available in English, as it is reported that the news agencies Xinhua and Toutiao and the media group Caixin are resorting to automated journalism, that it 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). However, the scarcity of material published in English prevents me from knowing more about the strategies these organisations develop. Little is known for instance about the "Media Brain" project that China's state agency Xinhua launched in 2018 (Xuequan, 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 artificial intelligence could be used to further disseminate propaganda coming from the Chinese Communist Party (see Ables, 2018).

2.3.2   Scholarship on 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 better spot the research orientations that characterise them (see Figure 4). For this purpose, I discerned two main fields of inquiry: first, the reach of automated news, which includes research on the perceptions of news readers (i.e., whether they rank it similarly to human-written content) as well as those on the wider repercussions of automated journalism (e.g., impacts on the legal and financial spheres); second, automated news in practice,

which encompasses studies looking into its functioning, its deployment within news organisations and its implications for media labour[15].
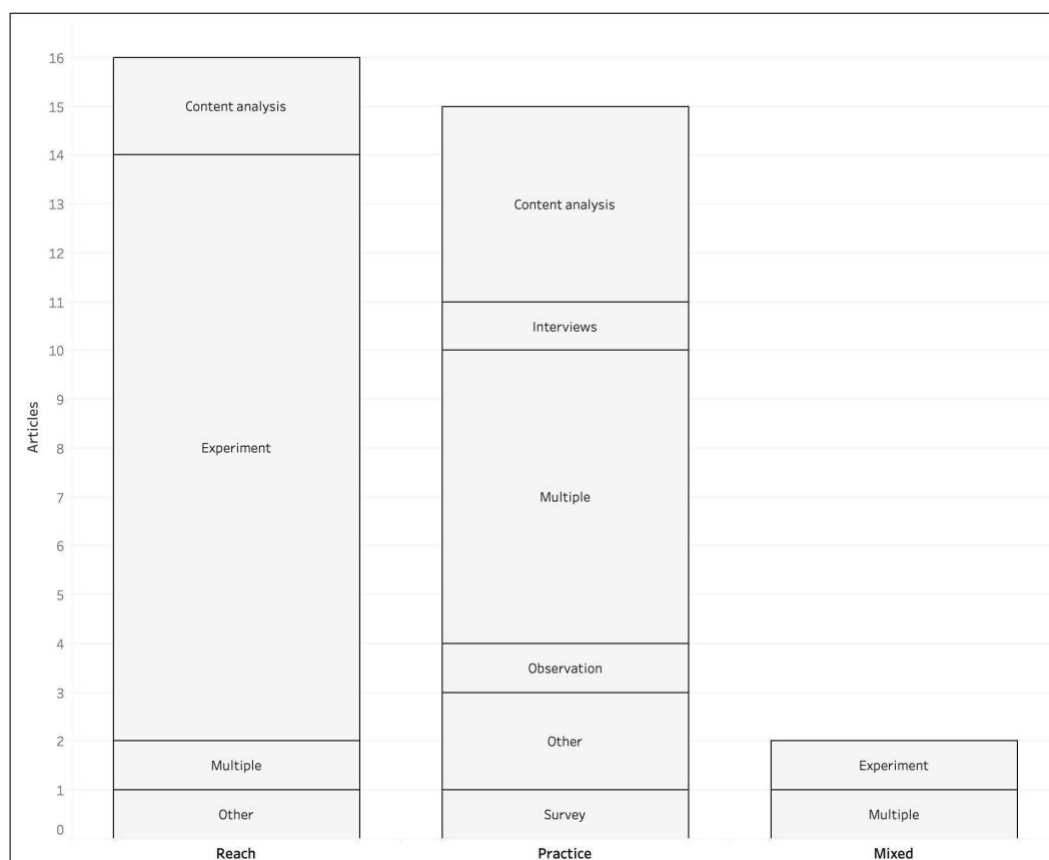
Figure 4: Methods used to study automated news. Methods employed for each field of inquiry that is featured in my corpus. Source: Danzon-Chambaud, 2021a.

Regarding the reach of automated news, I found that articles focusing on readers' perceptions were almost entirely constituted of experiments. These were conducted either solely (Clerwall, 2014; Graefe *et al.*, 2018; Haim & Graefe, 2017; Liu & Wei, 2019; Melin *et al.*, 2018; Tandoc, Lim & Wu, 2020; Waddell, 2018; Waddell, 2019a; Waddell, 2019b; Wölker & Powell, 2021; Wu, 2020; Zheng *et al.*, 2018) or in combination with other methods (Kim & Lee, 2019). In line with Graefe and Bohlken's findings (2020), they highlight in great part that readers evaluate the objectivity, trustworthiness and

---

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

credibility of automated news as being similar to human journalists', although when it comes to reading for pleasure they tend to prefer human-written news (Graefe *et al.*, 2018; Haim & Graefe, 2017; Melin *et al.*, 2018; Zheng *et al.*, 2018). In parallel, a growing stream of research is looking at combined authorship (i.e., humans and algorithms), which has demonstrated promising results so far (Tandoc, Lim & Wu, 2020; Waddell, 2019a; Waddell, 2019b; Wölker & Powell, 2021).

Besides readers' perceptions, other articles on the reach of automated news addressed the larger repercussions of the technology. Lewis, Sanders and Carmody (2019) and Díaz-Noci (2020) both resorted to content analysis to assess the legal impacts of automated journalism. They sustained that media organisations could potentially be condemned for negligence when "defamatory content slips through the cracks" (p. 15), while Díaz-Noci noted that human intervention in the creation of automated news could help news organisations secure copyright. In the business sphere, Blankespoor, deHaan and Zhu (2018) conducted a series of quantitative tests to analyse the market effects of automated news, which showed a correlation between automation of financial news and an increase in trading volumes for firms that were less covered prior to that. However, no impact on determining trade values was found.

In contrast to scholarship on the reach of automated news, those focused on practice involved many methods, most of the time combined together. I found that some technically-oriented studies looked at its functioning so as to demonstrate both its potential and limitations. As such, Caswell and Dörr (2018) performed a series of tests with a self-editing tool in order to create more complex template models: they demonstrated that reporting on news events this way made it possible for uncomplicated stories (e.g., car chases) to be automated, but not as far as more complicated ones (e.g., parliamentary proceedings) were concerned. In another technically-oriented study, Visvam Devadoss, Thirulokachander and Visvam Devadoss (2019) managed to create a fully operational NLG system able to draw on online content and on social media to feed a news website.

Another area that pertains to automated news in practice has to do with organisational impacts on newsrooms. In the area of sports journalism, two separate content analyses found, on the one hand, that commonalities between automated

and human-written news were more important than differences (Túñez-Lopez, Toural-Bran & Valdiviezo-Abad, 2019) and, on the other hand, that human intervention in the editorial process still remained significant (Rojas Torrijos, 2019). Regarding the use of "news bots" and "chatbots"[16], a digital ethnography established that news bots could help media outlets reach out to a niche and geo-specific audience, but sometimes lacked data transparency (Lokot & Diakopoulos, 2016), while a content analysis combined with interviews revealed that chatbots can be used to appeal to new audiences, but also need to be scrutinised and be made accountable so as to sustain public media values (Jones & Jones, 2019). Finally, another content analysis combined with interviews unveiled discrepancies in attribution policies, which led the authors to suggest that a more comprehensive framework needs to be developed (Montal & Reich, 2017).

A third and last domain that relates to practice looked at the perceived impacts of automated news on media labour. These studies included first-hand accounts gathered through interviews (Lindén, 2017a), surveys (Kim & Kim, 2017) or a combination of methods that, nonetheless, all involved interviews (Dörr, 2016; Kim & Kim, 2018; Thurman, Dörr & Kunert, 2017; Young & Hermida, 2015). Young and Hermida (2015) explored the launch of one of the very first automated news projects, the *Los Angeles Times*' Homicide Report, which covered every homicide in Los Angeles County in a programmatic fashion. They carried out, along with textual analysis, semi-structured interviews with the newspaper's Data Desk and Homicide Report teams, so as to see how a computational journalism mindset is developing within a legacy media organisation. They found, among others, that established actors like crime journalists were progressively losing influence to a "new class of computational journalist and non-human journalist" (Young & Hermida, 2015, p. 393), as programmers setting up automated news get to define what constitutes a homicide in the first place, leaving out complimentary details like the story's "human touch" to other newsroom staff. Lindén (2017a) relied on semi-structured interviews to probe news managers', data journalists' and other actors' (e.g., academics, consultants)

---

[16] 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.

impressions of journalistic logic being translated into code, as well as future developments in the area of automated news. He questioned the assumption that the latter would bring about job losses and argued instead that the "ideology of journalism" (see Deuze, 2005 in section 3.2.1)— in other words how journalists give meaning to their work—will be a strong mitigating factor, which already saw journalists retaining their jobs while navigating many other phases of automation. On their end, Thurman, Dörr and Kunert (2017) first carried out observations to study how journalists interact with a third-party self-editing tool, then proceeded to interviewing them to get their first-hand impressions of using the software. They concluded that the general impression is that automated journalism is faced with "fundamental limitations", like providing actual news context instead of just background information or understanding human nuances that makes for good reporting. Finally, Kim and Kim (2018) made use of exploratory interviews and of a psychology-inspired framework to evaluate the types of attitudes that journalists adopt when faced with automated news. A bit similarly to Powers (2012), they found that they either react through a "journalism's elitism" type of attitude, where they believe that automated news cannot supplant journalistic expertise, a "Frankenstein complex" that regards the technology more as a threat, or while adopting a more "rosy" view where journalists are keen to explore the new opportunities offered by automated news.

In parallel with these practitioner-oriented studies, additional scholarship use survey research to identify the dominating mindset among media executives (Kim & Kim, 2017)—who reportedly lean towards the implementation of automated news instead of hiring more journalists—or interviews with NLG providers in order to evaluate the market phase that automated news is in (Dörr, 2016). Two other scholarship engaged with qualitative content analysis to examine the perceived impacts of automated news at a more collective level. In the earliest study on automated journalism that was found, van Dalen (2012) suggested—after looking at 68 blog posts and newspaper articles—that journalists react to the introduction of the technology by emphasising their very human skills (e.g., creativity, personality). He also stressed that routine tasks could be assigned to automated news, so that humans can focus instead on more demanding formats (provided that newsrooms reassign them this way). Likewise—after having examined 63 pieces of media content,

websites and blogs—Carlson (2015) concluded that automated news could as much be used to alleviate or augment the work of media practitioners as to make them redundant. However, these studies investigating media labour on a broader scale were published at an early stage of automated news research, and can therefore be considered as exploratory.

2.3.3 Underexploited sociological frameworks

Finally, I examined my corpus to see to which extent theory was used—or not—and how this relates to the fields of inquiry described above (see Figure 5), this so in order to find meaningful insights that connect theoretical considerations to research on automated news. In most cases, though, no theory was used, which is especially true of studies focusing on practice: only three articles were actively making use of theory (Dörr, 2016; Kim & Kim, 2017; Kim & Kim, 2018) while two others were building on it, but did not strongly reflect this in the variables they used (Carlson, 2015; Lindén, 2017a). By contrast, half of the articles on the reach of automated news made an active use of theory as they dealt with readers' perceptions (Haim & Graefe, 2017; Liu & Wei, 2019; Tandoc, Lim & Wu, 2020; Waddell, 2018; Waddell, 2019a; Waddell, 2019b; Wu, 2020; Zheng *et al.*, 2018).
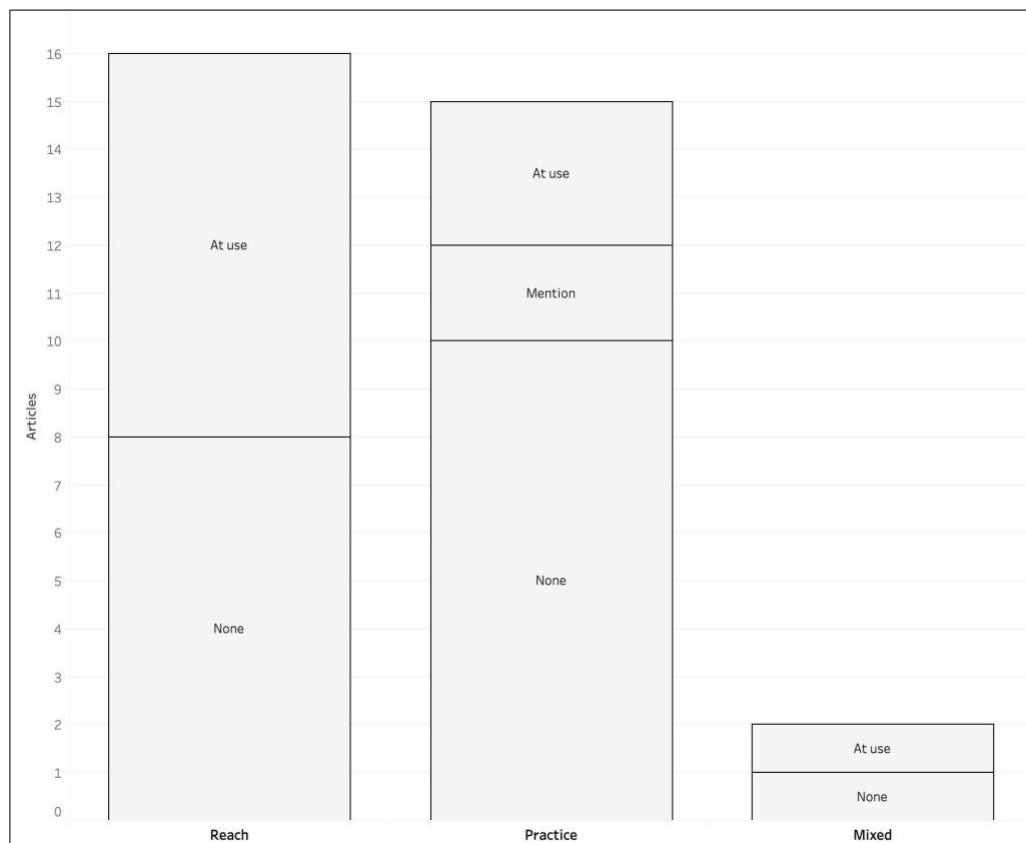
Figure 5: Use of theory to study automated news. Use of theory for each field of inquiry that is featured in my corpus. Source: Danzon-Chambaud, 2021a.

When comparing theory use with fields of inquiry (see Figure 6), I observed that scholarship on readers' perceptions resorted to psychological theories (i.e., expectancy violations theory in Liu & Wei, 2019; Waddell, 2018; Waddell, 2019a; Tandoc, Lim & Wu, 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)[17] while two studies on practice used sociological frameworks (i.e., institutionalism in Dörr, 2016; Kim & Kim, 2017) and another one a mixed sociological-psychological structure (i.e., innovation resistance theory and institutionalism in Kim & Kim, 2018)[18].

---

[17] One article employed a mixed sociological-psychological framework (i.e., high-context/low-context cultures and holistic/analytic thinking framework in Zheng *et al.*, 2018).

[18] One of the studies focusing both on readers' and journalists' perceptions of automated news also used a framework situated at the crossroad of sociology and psychology (In-group and out-group theory in Jung *et al.*, 2017).
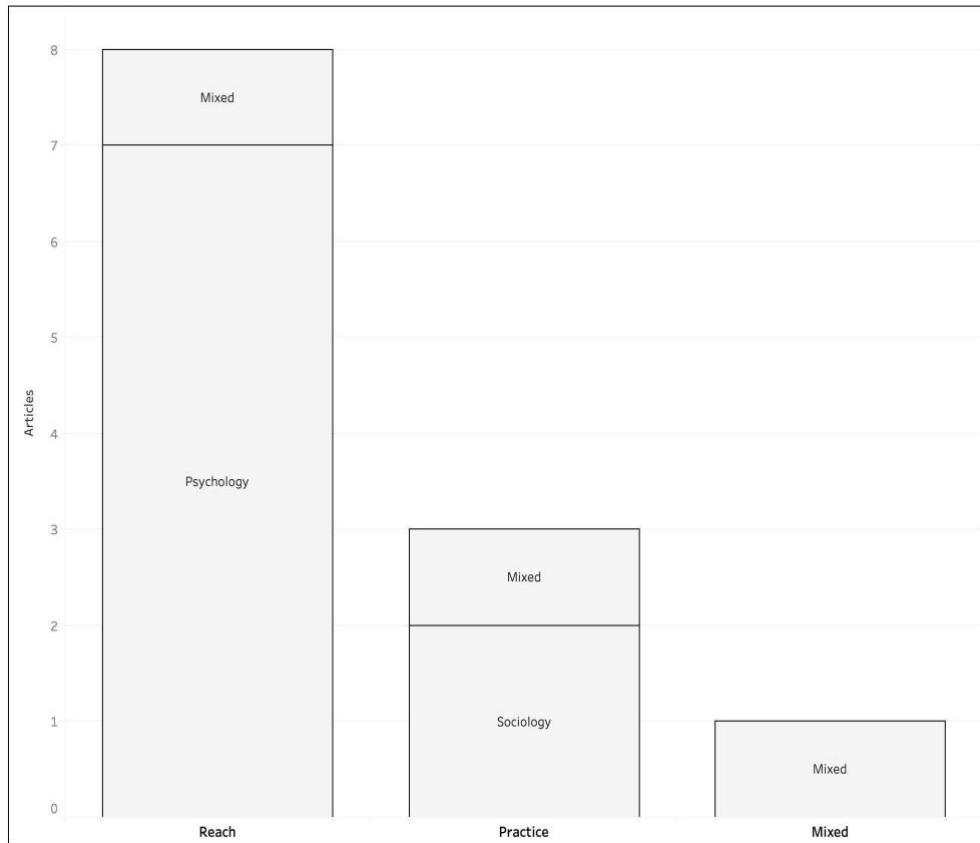
Figure 6: Theoretical backgrounds used to study automated news. Theoretical backgrounds for each field of inquiry that is featured in my corpus. Source: Danzon-Chambaud, 2021a.

To have a better idea of the sources behind these theories, I looked at the most-cited references throughout the entire corpus (see Table 2). First, I realised that the most-cited ones had to do with other empirical studies on automated news, which were sometimes already included in my corpus (Caswell & Dörr, 2018; Carlson, 2015; Clerwall, 2014; Dörr, 2016; Graefe *et al.*, 2018; Jung *et al.*, 2017; Lindén, 2017a; Lokot & Diakopoulos, 2016; Montal & Reich, 2017; Thurman, Dörr & Kunert, 2017; van Dalen, 2012; Waddell, 2018; Young & Hermida, 2015) or published in other formats like industry reports (Graefe, 2016), conference papers (van der Kaa & Krahmer, 2014) or book chapters (Lemelshtrich Latar, 2015). Additionally, other empirical material on algorithmic accountability (Diakopoulos, 2015), artificial intelligence for investigative reporting (Broussard, 2015), computational journalism (Karlsen & Stavelin, 2014;

Stavelin, 2013) as well as Powers' research (2012) on technologically specific forms of work (see section 1.1.1) also figured in this listing.

As for theoretically-oriented articles and book chapters, these were mostly concerned with exploring sociological aspects that relate to news making and digital and algorithmic transformation (Anderson, 2013; 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 sociological lenses they put forward remained largely unexploited in the articles I studied: even if a handful of studies on practice echoed some of these sociological suggestions (i.e., Anderson, 2013 and Napoli, 2014 in Dörr, 2016; Anderson, 2013 in Young & Hermida, 2015; Lewis & Westlund, 2015a in Thurman, Dörr & Kunert, 2017; Deuze, 2005 in Lindén, 2017a), none of them strongly reflected those in the variables they used.

Table 2. Most-cited references* in corpus. Source: Danzon-Chambaud, 2021a.

Number of citations	Publications
24	Clerwall (2014)
21	Carlson (2015)
20	van Dalen (2012)
17	Anderson (2013)
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, Dörr & Kunert (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 (2017a); 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.

Besides, only a few theoretically-oriented articles advised to turn to well-established theories to look at the way algorithms are transforming journalism. Dörr and Hollnbuchner (2017) recommended using traditional theories of ethics (i.e., deontology, utilitarianism, virtue ethics and contractualism) while Napoli (2014) referred to Institutionalism to emphasise how a social constructivist approach and isomorphic processes could help understand algorithmic media consumption and production. Together with Institutionalism, Anderson (2013) advised resorting to Bourdieu's Field theory to bring, as written earlier, "a vector of power dynamics" to the field of technological innovation. Institutionalism is also actively used in a few studies on professionals[19]: Kim and Kim (2017) used the concept of institutional isomorphism to evaluate whether media executives decided to implement automated news based on trends followed by other organisations, and in their study on journalists' attitudes towards the adoption of automated news (2018) they used the Institutionalism-inspired concept of institutional change, according to which structural transformation is likely to bring uncertainty and therefore stress among workers. Finally, the Bourdieusian lenses suggested by Anderson remained unexplored in the publications I systematically retrieved, thus indicating a research gap in this area. Outside of my corpus, though, it is important to stress that Field theory has been used at length in Wu, Tandoc and Salmon's (2019b) examination of algorithmic automation within newsrooms (to be further discussed in section 3.3.2).

Finally, one last type of publications that are visible in my listing are contributions focusing on readers' perceptions and evaluation of credibility, whether these concerned communication content at large (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 messages as such (Appelman & Sundar, 2016). Contrarily to sociological frameworks, these psychologically-inspired studies were largely operationalised as variables in articles investigating readers' perceptions (i.e., 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, 2019b and Wu, 2020; Sundar & Nass, 2001 in Zheng,

---

[19] Dörr (2016) also made active use of Institutionalism-inspired concepts, but his work is based on publications that were only available in German, thus preventing me from having a clearer idea of the concepts he used.

Zhong & Yang, 2018; Meyer, 1988 in Liu & Wei, 2018, Wölker & Powell, 2021, Tandoc, Lim & Wu, 2020; Wu, 2020; Appelman & Sundar, 2016 in Waddell, 2018, 2019a, 2019b and Liu & Wei, 2018).

2.3.4 Additional scholarship

In addition to the systematic literature review above, I am also including here a more flexible account of material that fell outside my search criteria, such as scholarship that did not figure in the databases I looked at or that was published after mid-2020, as well as other formats like non-academic article content. Some of these formats included books and chapters, one of which portrayed automated journalism in quite a dystopian manner: in his book chapter, Lemelshtrich Latar went as far as to advance (2018, p. 29) that "within 5–10 years, the majority of all journalistic text stories will be written by robots" and that it represented a "huge threat" to press freedom and public interest journalism. Naturally, such views need to be mitigated as Lindén and Dierickx (2019, p. 155) did as they introduced the metaphor of automated news as a "washing machine", instead of the habitual depiction of a "robot journalist" taking over journalistic jobs. As they framed it (*ibid.*, p. 155), journalists are actually "in charge of sorting the dirty laundry (i.e. data), of choosing the right washing programme (i.e. what is the narrative) and pushing the button (i.e. deciding on the medium and timing of the story)". By contrast to Lemelshtrich Latar, Diakopoulos' book (2019) on news automation offers a much more nuanced perspective, which highlights the advantages of using the technology—like speed, publication at scale and personalisation of news content—but also limitations that dealt for instance with data dependency and data quality issues.

As for industry reports, an important piece to mention is Graefe's *Guide to Automated Journalism* (2016), which underlined the opportunities and limitations of automated news as well, but also transparency issues, especially with regard to errors or when personalisation is at play. Another report edited by Lindén and Tuulonen (2019) emphasises for its part that template writing should come with editorial oversight, and that extending automated news to domains where possible outcomes are not predictable enough calls for additional caution (see the difficulties to automate parliamentary proceedings in Caswell & Dörr, 2018). It also sheds light on

data privacy issues that relate to personalisation, even more so considering the European Union's General Data Protection Regulation. In a last industry report targeted, this time, at media councils, Haapanen (2020) argued that these bodies should be prepared to lodge complaints that deal with automated news usage, especially as news personalisation is involved. He also advances that, should media councils not endorse this role, other institutions like national legislators, the European Union or even platforms companies will, thus jeopardising press freedom.

Looking this time at academic articles and conference papers, these can be split, again, between studies on reach and on practice. Publications focused on reach of made use of yet another series of experiments that assessed readers' perceptions of automated news. In a meta-analysis that examined the results of these experiments in 12 papers, Graefe and Bohlken (2020) came to a conclusion close to mine, namely that there is no difference in readers' perceptions of credibility in automated news content and in human-written news, and that human-written news are better evaluated in terms of readability. They also observed that readers tend to give a better score to a story when told it is written by a human journalist, thus challenging the view that readers would prefer automated news as a solution free of bias (see also Gillespie, 2014), and thus potentially encouraging media organisations not to reveal that the story has been generated using algorithms, which raises ethical flags of all sorts.

Other experiments on news readers brought forth additional perceptions from groups not previously studied, like media practitioners' own evaluations in van der Kaa and Krahmer (2014), who demonstrated that journalists evaluated automated news less positively than other readers with regard to trustworthiness. There were also experiments that investigated the types of heuristics at play when reading automated news, as in Jia and Johnson's (2021) examination of readers' predispositions to select news content based on personal values. They established that attributing the story to an algorithm could potentially reduce this effect when it came to heavily politicised subjects, like gun-rights stories. However, in stories appearing to be rather neutral, Jia and Johnson found no difference in perceptions of credibility between automated news content and human-written content.

Scholarship on practice, on their end, fell within two categories: reflections on conceptual notions around automated news and empirical accounts documenting its deployment within newsrooms. Looking at conceptual notions first, Dierickx (2021) suggested using software studies—a new research stream that tries to analyse how software fits within a broader societal context—together with journalism studies as the intersection of editorial decision-making with NLG processes (*ibid.*, p. 5) "could be seen as a technological black box that embeds a journalistic black box". Additionally, software studies could bring added value in that it helps seeing journalism as a process that can be deconstructed, and as such can be used to interrogate the mutual shaping relationship between human-made choices and technology. Leppänen, Tuulonen and Sirén-Heikel (2020) also looked at the broader societal context behind automated news while identifying the kinds of biases it may potentially trigger: they singled out, first, biases that relate to content selection and, second, biases that relate to language choice. The authors suggest nonetheless that biases in automated journalism are unavoidable, as both template-based and advanced machine learning techniques depend on someone's or a team's set up, which is ultimately influenced by structural and organisational context.

In addition to these conceptual notions, other scholarship examined automated news in practice using research interviews so as to evaluate its perceived impacts on media labour. To a certain extent, these results can be interpreted using Powers' three categories of how journalists react to new technological capabilities (see section 1.1.1): in Lindén (2017b), the adoption of automated journalism at the Associated Press corresponds to Powers' idea of an extension of occupational norms, but at Local Labs in Chicago it rather matches the notion of "threat" as it translated into job losses; Lindén also reported that new forms of work centred around computational thinking saw the light of the day at ProPublica, which echoes Powers' category of new forms of work that can be used to reinvent occupational norms. Powers' idea of an extension of occupational work is also visible in other scholarship on automated news in practice: after conducting interviews with American and European media representatives, Sirén-Heikel *et al.* (2019, p. 61) concluded that "automation can be seen as part of the same continuum that replaced the pen with first a typewriter and then a computer"; likewise, Schapals and Porlezza (2020) found that, in German newsrooms, automated journalism is rather seen as a way to advance

media practitioners' own work than posing a threat to it, a conclusion that is shared in Kunert's (2020) analysis of the relationship between data providers, software companies and media organisations when creating automated sports stories in Germany.

2.4 RESEARCH QUESTIONS

In this chapter, I went from identifying the wider technological and journalistic context behind the development of automated news to giving a detailed account of scholarship published on it. In my systematic analysis, I noticed that there is a need to take a better look at the perceived impacts of automated news on media labour: unlike readers who see no differences in objectivity, trustworthiness and credibility between automated news and human-written content—but prefer the latter in terms of readability (see also Graefe & Bohlken, 2020)—there is no similarly clear pattern as to how it affects practitioners. To document this with a critical eye, the use of well-established theories, especially in sociology, would be particularly relevant as my review showed a lack of practice-oriented research resorting to these types of lenses, which would be all the more important to use when reflecting on this changing landscape at a higher level. Finally, as stressed in the Introduction, it would be worthwhile to link these considerations to the overall relationship between journalism and technology. In light of all these interrogations—which can be considered as this study's research gap—my research questions therefore go as follows:

- RQ1. What are media practitioners' perceived impacts of automated news on the work they do?

- RQ2. What do these considerations entail for journalism practice and for journalism as a whole?

- RQ3. How can these reflections advance our understanding of the relationship between journalism and technology?

In the next chapter, I will introduce two theoretical approaches, Actor-network theory and Bourdieu's Field theory, which I deemed to be best suited to investigating the deployment of automated news within newsrooms and to analysing power dynamics that result from it, this so despite the apparent ontological differences between the two.

3   Theoretical considerations

In his study of practice theory, Ryfe (2018) observed that Actor-network theory, Institutionalism (which includes here Gidden's Structuration theory, but could also arguably be extended to New Institutionalism) and Bourdieu's Field theory have caught scholars' attention when it comes to examining the sociology of news. In the corpus of news production literature he examined, he found for instance (p. 224) that essays on Actor-network theory "almost exclusively discuss the theme of technology as a social actor and its uses for studying journalistic innovation", while "essays on field theory tend to be oriented toward comparative research across journalistic fields[20], and those on institutionalism toward organizational studies". Having said that, I will focus in this dissertation on two of these frameworks—Actor-network theory and Field theory—which I believe are the most suited to, first, studying the many ways automated news is deployed within newsrooms (i.e., Actor-network theory), second, investigating the new power plays that result from it (i.e., Field theory), especially given the different types of external influences that lie behind journalist's use of technology (see section 2.2). As I will explain next, Actor-network theory and Field theory present fundamental ontological differences; yet I believe that, despite these, each can bring valuable insights when looking at them separately to see where they intersect.

3.1   Actor-network theory

Originally developed in the 1980s by science and technology scholars Michel Callon, Bruno Latour and John Law at the École des Mines de Paris (Buchanan, 2010), Actor-network theory (i.e., ANT) essentially revisits sociology from a "bottom-up" perspective, following in that sense a social construction of technology perspective which posits that scientific facts cannot be seen as truths of their own, but also as partially constructed knowledge (Bijker, 2008): for instance, this could be the case

---

[20] For instance, this type of comparative analysis can span across countries, as in Benson's work (2013) on French and American journalistic fields.

when research orientations are being prioritised over others or when scientific discoveries are being made thanks to technical means available at the time (e.g., using microscopes to be able to spot bacteria). It is against this backdrop that Latour questions what he calls (2005, p. 35) the "sociology of the social", in other words the use of well-established sociological frameworks to analyse society, which according to him are prone to "social inertia". Instead, he suggests to go for a "sociology of associations" that is designed to follow the actors' trails:

> The choice is thus clear: either we follow social theorists and begin our travel by setting up at the start which kind of group and level of analysis we will focus on, or we follow the actors' own ways and begin our travels by the traces left behind by their activity of forming and dismantling groups.
>
> (Latour, 2005, p. 29)

Thus, ANT encourages researchers to start tracing connections from scratch as they leave aside any pre-constructed sociological explanations[21], even though these may act as "companion concepts" that can be encountered at a later stage (Winthereik, 2020).

3.1.1   *Human*, *non-human* and *translations*

One key aspect of ANT is that it takes into account a "rich bestiary of significant actors" (Clark, 2020, p. 160)—or rather *actants* (Crawford, 2005; Blok, 2019)—which involves *human* and *non-human* elements that can be as diverse as (Michael, 2017a, p. 5) "mundane objects, exotic technologies, texts of all sorts, nonhuman environments and animals". The term "actor-network" in itself speaks to the idea that every actor and all of its attributes—such as thinking, writing or loving for humans—are never entirely cut out from each other, thereby creating a "web of relations" that stretches "both within and beyond the body" and across which action

---

[21] To a certain extent, ANT can be linked to Grounded theory, a process that can be broadly described as developing "tentative theoretical propositions" from the get-go (Scott, 2009, p. 448), or in other words "generating theory from data" (Scheufele, 2008). However, ANT differs from Grounded theory, in that it is rooted in a more social construction of technology perspective.

is distributed (Primo, 2019, p. 2; Law, 1992, p. 384). To better understand this, Law uses the following metaphor about himself (1992, pp. 383–384): "If you took away my computer, my colleagues, my office, my books, my desk, my telephone I wouldn't be a sociologist writing papers, delivering lectures, and producing "knowledge." I'd be something quite other—and the same is true for all of us." As such, ANT is therefore well suited to studying change in practice (Plesner, 2009); in the case of journalism, it helps account for all the "tools of the trade" (e.g., web searches, databases, smartphones) that make it as it is today, and can be used to document journalistic innovation, including automated news (Primo, 2019, p. 2). Besides, ANT can also be leveraged to study newsroom infrastructure: for instance, Mari (2021) used it so as to get a sense of the impacts of air conditioning on news work.

Another marker of ANT is the concept of *translation*: not to be confused with language translation, ANT's *translations* rather speaks to a phenomenon whereby *heterogeneous* entities (Law, 1984) or *actants* come together to form an actor-network, thus potentially disengaging themselves from other structures they belong to[22]. As illustrated in Figure 7—which depicts how a scallop species (*Pecten Maximus*), local fishermen, the scientific community and a team of biologists engaged into forming an actor-network whose goal is to replenish scallops beds in the Saint-Brieuc bay, in France—each of these entities *translate* their interests so that they pass them through an *obligatory point of passage*—a common objective—which makes the network hold; in this case, it is the biologists' novel research programme, who then become *spokespersons* for the group. As Callon (1984, p. 224) put it: "Translation is the mechanism by which the social and natural worlds progressively take form. The result is a situation in which certain entities control others." That being said, for the actor-network to be able to last in time, successful *enrolment*, or (*ibid.*, p. 211) "the device by which a set of interrelated roles is defined and attributed to actors who accept them", needs to be sustained, making it a structure where relational power is always up for negotiations (Michael, 2017a). If robust enough, though, it may give rise to a *macro actor* that is able to restructure society as whole (Czarniawska, 2016;

---

[22] The use of the term *translation* in ANT can be better understood when looking at its Latin roots: here, translation rather refers to *trans-latio*, which makes reference to a change in location (Czarniawska, 2016).

Cooren, 2019), as in Latour's (1988) example of how Louis Pasteur and his scientific discoveries—which constituted an actor-network of their own—ended up reshaping France's agricultural and societal landscapes (Blok, 2013; 2019).
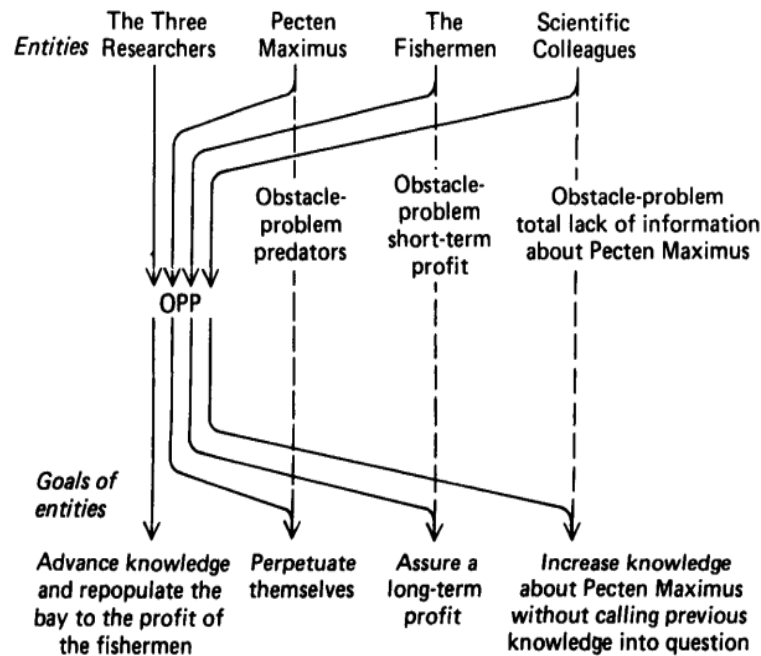
Figure 7: ANT's *translation* process. This schema represents how three researchers positioned themselves as *spokespersons* in a farming experiment aiming at replenishing scallops beds (*Pecten Maximus*) in the Saint-Brieuc bay, in France: their research program became the obligatory point of passage (OPP) that other actors need to pass through in order to *translate* their own interests. Source: Callon, 1984.

3.1.2 *Intermediaries* and *mediators*

In ANT, the process through which entities undergo *translations* and become stabilised enough to form an actor-network comes under terms like *simplification, black-boxing* or *punctualisation* (Crawford, 2005; Michael, 2017a). Once firmly established, they become set structures passing on the same type of predictable output based on any given input, and are also able to act at distance: such was the case, for instance, of different navigational elements (e.g., winds, star positions…) being *translated* and ultimately *black-boxed* into a vessel-actor-network, the

Portuguese carrack, which came to operate while making journeys at sea as much predictable as possible (e.g., planned routes, standardised ways of observing stars to be able to self-locate), thus asserting Lisbon's *long-distance control* over overseas territories (Law, 1984). In ANT, stabilised actor-networks like these are known as *immutable mobiles* or *inscriptions* when they concern textual or graphical material (Latour, 2005; Crawford, 2005; Nikolova, 2008; Hassard & Alcadipani, 2010; Michael, 2017b), and may come, in the case of technological artefacts, with particular *scripts* or "rules of use" that condition human behaviour (Akrich, 1992; Michael, 2017a). Moreover, Latour (2005) distinguishes between two modes through which action can be distributed: first, as *intermediaries*, where meaning is maintained and where outputs can be predicted by inputs; second, as *mediators*, where meaning is changed and where inputs are never a good predicator of outputs. Drawing on Callon (1991) and Latour (1992), Sayes (2014, p. 138) specifies that "an intermediary is a placeholder in the sense in which it merely does what anything else in its position would do", whereas "a mediator is something *more* than this": in the case of non-human elements, it is "seen as adding something to a chain of interaction or an association". To explain these specifics, Latour gives the following example:

> A properly functioning computer could be taken as a good case of a complicated intermediary while a banal conversation may become a terribly complex chain of mediators where passions, opinions, and attitudes bifurcate at every turn. But if it breaks down, a computer may turn into a horrendously complex mediator while a highly sophisticated panel during an academic conference may become a perfectly predictable and uneventful intermediary in rubber stamping a decision made elsewhere.

> (Latour, 2005, p. 39)

Analytically speaking, ANT can be used to look inside all of these *black-boxes* that have formed over time (Primo, 2019), or to put it another way to uncover the "Big Leviathan" that a *macro actor* like society is made of (Callon & Latour, 1981). Additionally, Latour (2005, p. 108) also suggests to consider that "*there exist translations between mediators that may generate traceable associations*": ANT researchers should then focus on establishing these connections in order to "follow the actors' trails" (Primo, 2019). As Latour expressed it:

> To put it very simply: A good ANT account is a narrative or a description or a proposition where all the actors *do something* and don't just sit there. Instead of simply transporting effects without transforming them, each of the points in the text may become a bifurcation, an event, or the origin of a new translation. As soon as actors are treated not as intermediaries but as mediators, they render the movement of the social visible to the reader.

(Latour, 2005, p. 128)

On a practical level, ANT can facilitate a bottom-up investigation of any given organisation while shedding light on all the interactions and textual exchanges that are taking place there, with a special attention being given to *translations* like (Cooren, 2019, p. 2) "the filling in of forms, the designing of plans, the mobilization of machines, all partaking in the unfolding and tracing of an organized and collective activity". Following Latour's recommendation to "follow the *mediators*' trails", ANT can be used to map out a state of flux and to help pin down the general direction that a social phenomenon is taking, thus contributing to determining power relationships between entities.

3.1.3   ANT in media research

In media and communication research, ANT can be used to investigate the introduction of a new technological artefact in a well-established actor-network, especially as it is faced with resistance: indeed, as the new technology is being embedded into the actor-network with its own intended meaning, a series of mutual *translations* happens, resulting in new power relationships. For example, while studying the deployment of a Personal Digital Production system (i.e., PDP) at the BBC—which allowed editorial staff to film and edit videos on their own—Hemmingway (2005, p. 25) found that a "initial rejections of PDP came from those people on the network unable to organise resources until PDP operators were also internalised and socialised within the network".

In parallel, ANT can be employed to establish power relationships between entities that already exist in a given actor-network, as in Schmitz Weiss and

Domingo's account of innovation in online newsrooms (2010). They observed (*ibid.,* p. 1063) that "the obligatory point of passage of the production team as translator of journalistic needs into technological developments hindered opportunities for innovative ideas to flourish", as "breaking news reporters felt their ideas were neglected, and web developers limited themselves to following instructions from the online editor". According to them (*ibid.,* p. 1068), this made online journalists feel "powerless in the decision-making process" while technologists viewed their colleagues' needs as a lack of skills. In the case of Internet and digital technologies being brought into Greek newsrooms, Spyridou *et al.* (2013) looked at interactions and power relationships between well-established actors (i.e., journalists), new actors (i.e., technological tools, convergence and participation) and former *intermediaries* with an increasingly important role to play (i.e., audiences), in order to find out, among others, whether the dominant journalistic culture adapts to or contributes to shaping new technological affordances. They eventually conclude (*ibid.,* p. 93) that these changes "tend to get normalized or 'rationalized' through the values and norms of the dominant journalistic culture".

3.2   BOURDIEU'S FIELD THEORY

Contrarily to ANT's focus on individual *actants* that gradually become entangled within a network, Bourdieu's Field theory[23] rather takes a "mezzo-level" scope of analysis as a starting point, which is able—when applied to journalism—to account for all the dimensions located (Benson, 2006, p. 199) "between the individual news organization and the society as a whole". In doing so, Field theory aims at locating the internal tensions occurring within each of these microcosms, or "fields" (Bourdieu, 2005a). Although Bourdieu's model was initially developed to analyse artistic behaviour (Bourdieu, 1992) and was later used to examine the influence of televised news (Bourdieu, 1996), Field theory can prove to be useful when scrutinising the impacts of digital media on journalism (Benson & Neveu, 2005).

---

[23] The overview on Bourdieu's Field theory that is detailed here reproduces some of the contextual information used in a methodological article I wrote with my supervisor (see Danzon-Chambaud & Cornia, 2021b).

3.2.1 A world of fields, full of struggles

In Field theory, Bourdieu envisions the social world as a myriad of *fields*, continuously exposed to an struggle between two major forces: on the one hand, a form of power that arises from *economic* capital, or in other words "money or assets that can be turned into money" and, on the other hand, a form of strength made of *cultural* capital, which we can understand as a set of unique abilities that include "educational credentials, technical expertise, general knowledge, verbal abilities, and artistic sensibilities" (Benson & Neveu, 2005, p. 4). Other forms of capital also come into play (Jenkins, 2005), like *symbolic* capital (e.g., honour and reputation) or *social* capital (i.e, networks and relationships). With regard to the journalistic field, *journalistic* capital constitutes the very type of cultural capital that is relevant to this field (Schultz, 2007), which sometimes encompasses *social* and *symbolic* capital as well (Meyen & Riesmeyer, 2012). As for *technological* capital, it can be viewed as a distinct form of capital—as in Romele's (2021, p. 495) vision of technology as a form of strength that social actors or groups can wield so to act as an authority and consequently rearrange "a technologically mediated social world" according to their needs—or as being part of cultural capital altogether, Bourdieu (2005b, p. 194) describing it as a set of "procedures, aptitudes, routines and unique and coherent know-how, capable of reducing expenditure in labour or [financial means] or increasing its yield". All these forms of capital are often field-specific, but can sometimes be converted into mainstream forms of capitals that have value across fields. As Crossley (2021) explained:

> Publishing in a high-ranking sociology journal only really has any value or meaning for a sociologist, for example, and extensive vascularity is similarly only of value to bodybuilders. [However,] a good publication may help the sociologist achieve promotion, for example, which will increase their income, and vascularity may help the bodybuilder to win a competition with a big cash prize. Moreover, successful bodybuilders and (less often) sociologists might achieve a degree of celebrity, with the further opportunities this opens up.

> (Crossley, 2021, p. 2)

In addition to these, one core tenet of Field theory is that, within each field, the ongoing competition between economic and cultural capitals translates into two poles, which can be found under different wordings depending on the field (Bourdieu, 2013), but are known as the *heteronomous* and *autonomous* poles in journalism. According to Bourdieu (2005a), the *heteronomous* pole reflects a type of journalism exposed to external influences, mostly political and economic, as illustrated for instance by the influence that advertisers exert on commercial television news. At the other end of the spectrum, the *autonomous* pole is considered to be a manifestation of an independent form of journalism, which Bourdieu perceives as being the "purest", based on the assumption that it would be exempted from external pressures. Print journalists with editorial independence that determine the news agenda for the day and journalists being awarded the Pulitzer Prize each year could, in a sense, be representative of this *autonomous* pole (Bourdieu, 2005a; Benson & Neveu, 2005). At the time he developed these ideas, Bourdieu described the journalistic field as being increasingly *heteronomous*, mostly because of the influence that television news held on other forms of journalism. According to him (Bourdieu, 2005a), the pressure put on by advertisers through audience ratings was further reinforced by the existence of precarious labour and employment, which made censorship easier through political and economic control.

In addition to the confrontation between these two poles, Field theory also introduces concepts that relate to an individual's behaviour within the field: hence, the *doxa* reflects the "universe of tacit presuppositions that we accept as the natives of a certain society" (Bourdieu, 2005a, p. 37), the *habitus* assumes that "individuals' predispositions, assumptions, judgments, and behaviors are the result of a long-term process of socialization", while the *illusio* represents "an agent's emotional and cognitive 'investment' in the stakes involved" (Benson & Neveu, 2005, p. 3). In the journalistic field, the *doxa* can be conceptualised as the "rules of the game" (Tandoc & Jenkins, 2017), the *habitus* as a "feel for the daily news game" (Schultz, 2007) and the *illusio* as a belief that this game is "worth playing" (Benson & Neveu, 2005). However, articulating the very specifics of these "rules" or *doxa* is up for interpretation. To provide a common understanding of what can be perceived as the journalistic *doxa*, I rely on Deuze's comprehension of journalism ideology—or "how

journalists give meaning to their newswork" (2005, p. 444). Often used as a normative model in media studies (see Lindén, 2017a; Usher, 2017), Deuze introduces five *ideal-typical values* that he believes to be representative of journalism's ideology: *public service*, characterised by a strong sense of public mission and by intending to serve the people, most notably through a "watchdog" style of reporting; *objectivity*, which speaks to a sense of impartiality, fairness and professional detachment; *autonomy*, which relates to journalists' freedom to tell the stories they want without external forms of pressures, constraints or influences; *immediacy*, which refers to the speed of breaking news, but also to the 24/7 news cycle (see section 1.1.1); and *ethics*, which can be understood as individual behaviours being regulated by professional standards.

Finally, Bourdieu (1997) also evokes a situation of *hysteresis* or "Don Quixote effect" where individuals "judge and act today according to dispositions previously acquired under quite different social conditions" (Benson & Neveu, 2005, p. 10), thus resulting in their *habitus* being out of touch with a new order (Wu, Tandoc, and Salmon, 2019b). This could be, for instance, farmers in Southern France being unable to court women using the *habitus* that had prevailed until the growing influence of urbanisation (Bourdieu, 2002). According to him, such an *hysteresis* effect can happen when a field is undergoing a major crisis, thus profoundly changing its regularities and even its rules.

To illustrate how all of these hidden mechanisms actually play out, Bourdieu resorted to metaphors like those of a game or of a marketplace (Jenkins, 2005). In one of his preliminary works on the notion of "field", he used an auction as a metaphor, even though some of the precise terms were coined at a later date: in an auction, he argued, tentative buyers have a shared understanding of the stakes at play (i.e, the *illusio* that acquiring goods this way is worth it), they agree on legitimate ways of acquisition (e.g., stealing these items is not considered to be part of the auction's *doxa*) and, finally, they have limited amounts of resources that are not only made of economic capital (e.g., they also have cultural capital that helps them see in an old drawer a splendid work of art, once repaired by a craftsman). As far of the field of journalism is concerned, one may envision news workers calling on their many types of *capital* (e.g., cultural, symbolic, social, or simply journalistic) so as to get news as fresh as possible (i.e., the *illusio*) while abiding to a set of common principles and

practices (i.e., the *doxa*). like those defined in professional codes of conduct, but also unspoken ones like the journalistic *gut feeling* (see Schultz, 2007). That said, this type of struggle that is proper to the journalistic field may lead to homogeneity as well, as the influence of commercialism could bring "uniformity, censorship and even conservatism" among the news media. As Bourdieu explained with regard to the French media landscape:

> One very simple example: the battle between the three French weekly news magazines, *Le Nouvel Observateur*, *L'Express* and *Le Point*, results in their being undistinguishable. To a large extent this is because the competitive struggle between them, which leads them to an obsessive pursuit of difference, of priority and so on, tends not to differentiate them but to bring them together. They steal each other's front page stories, editorials, and subjects.
>
> (Bourdieu, 2005a, p. 44)

3.2.2 Bourdieu and *differentiation/de-differentiation* concepts

These observations made by Bourdieu are not dissimilar to a process of *de-differentiation*[24] that has been happening in Western European and North American media history, which is a core concept of Hallin and Mancini's media system typology (2004). Hallin and Mancini distinguish three types of media systems, based on their analysis of a set of dimensions that range from the structure of media markets to professionalisation and the role of State. These are: the "Mediterranean" or "Polarised Pluralist Model", which includes countries such as France, Spain and Italy and is characterised, among others, by a low level of journalistic professionalisation—not dissimilar to political activism—and by strong connections with the State given delayed liberalisation in these countries (even if commercial influences have progressively grown in importance); the "North/Central European" or "Democratic Corporatist Model", which concerns countries like the Nordics, Germany and Switzerland and where the media are considered social institutions that need to be

---

[24] In sociology, *de-differentiation* can be understood as a process that runs counter to *differentiation*, which assimilates modernity with "dividing society into stratified subsystems with specific specializations" (West, 2021). By contrast, *de-differentiation* implies that these specialised structures return to a more homogeneous form (*ibid.*).

protected by the State due to the pluralistic and consensus-based nature of these democracies, but still have a high degree of commercialisation and journalistic professionalisation; and the "North Atlantic" or "Liberal Model", which extends to countries like Canada, the United States and the United Kingdom, where commercialisation and journalistic professionalisation are relatively high and the role of the State moderated, even if sometimes commercial influences can circumscribe journalistic independence.

Following the logic of *differentiation*, Hallin and Mancini argued that the North Atlantic model of journalism sits the furthest away from social and political structures while the Mediterranean model presents strong ties between media and politics, which appear as two fields or sectors that often overlap. Finally, the North/Central European model is often situated somewhat in-between these two systems. They also observed that a process of *de-differentation* driven by market forces seems to be steering the Mediterranean and North/Central European models further away from socio-political influences to bring them closer to the types of commercial values found in the North Atlantic model, resulting in making these media systems more homogenous, even if differences in national political systems prevent them from being totally similar. To a degree, this is similar to the shift observed from political influences in the age of print and television news to audience and advertising ones in the age of digital, which now extends to a new form of technological pervasiveness (see section 2.2). That said, recent scholarship suggested that—even though Hallin and Mancini remains relevant to analyse media and political developments today—there is evidence that their existing models of journalism could also be converging towards a hybrid "Polarized Liberal" system, which started to be discussed in the wake of the 2016 presidential election in the United States and following Trump's presidency (Nechushtai, 2018).

Going back to Bourdieu's argument, this is where the reinforcement of economic forms of power over non-profit ones has translated into a hyper-commercialisation of the journalistic field that, in the end, is making the *heteronomous* pole more prominent. In newsrooms, this is visible in the erosion of the "old wall" that separates editorial from business (Benson, 2006; Cornia, Sehl, & Nielsen, 2020); in media content (Hallin & Mancini, 2004), this can be seen in increasingly blurry lines

between news and entertainment (i.e., "infotainment"), advertising and entertainment (i.e., product placement) and news and advertising (i.e., cross-promotion of products that belong to the same media conglomerate). Besides, Hallin and Mancini argue that a broader process of *de-differentiation* could be at play, based on Bourdieu's assumption (2005a, p. 44) that, because of the advantageous position the field of journalism has *vis-à-vis* other fields, "this journalism, increasingly dominated by commercial values, is expanding its domination over other fields"—or to put it differently "is tending to strengthen the most heteronomous zone in each of the fields—scientific, legal, philosophical, etc". This needs to be mitigated, though, by the possibility that the shift to commercialisation actually reinforces political influences, as commercial media can also reflect a class bias that generally leans towards the political right (see Murdock & Golding, 1977; Westergaard, 1977; Curran, 1979; as exposed in Hallin & Mancini, 2004). Such instances have originally been documented in countries with a high level of liberalisation, like the United Kingdom, but are now visible in other countries that have followed the same commercial logic.

3.2.3 Field theory in media research

In media studies, Bourdieu Field theory's concepts have been operationalised, among others, to look at the various forms of capital that are at stake within the journalistic field or in its subfields, and establish how they are being distributed. Siapera and Spyridou (2012) focused on how the introduction of online journalism within the field contributed to either generating, amplifying or depleting the economic, cultural, social and symbolic forms of capital that already existed there. Their analysis led to mixed results, for instance when observing that online journalism's economic capital remained lower than those of print and broadcast journalism, even though it was on the rise as opposed to print journalism's diminishing revenues. Another example relates to social capital, with online journalists having a harder time getting access to formal membership groups (e.g., journalistic unions), but being able to develop connections with news workers, sources and audiences through the use of social media. English (2016) used a similar logic when mapping out, first, the subfield of sports journalism within the broader journalistic field, then the position of individual journalists, news organisations (with a focus on broadsheet/quality titles), types of

publication and content in this subfield, as well as the nations that they belong to (i.e., Australia, India and the United Kingdom). To do this, he determined the position of each actor and entity while relying on a popular grid that is used in Bourdieusian studies (see Vandenberghe & Peters, 2019), which features a horizontal axis spanning from high journalistic[25] capital to high economic capital, and a vertical axis showing the amount of total capital they hold (see Figure 8), thus showing whether they belong to a dominated or dominating group. English's analysis demonstrated for instance that sports journalism has more "total capital" than world news in the broader field of journalism and that, in the subfield of sports journalism, senior writers and specialist reporters sit in the most influential position.

---

[25] To oppose *economic* capital, English uses *journalistic* capital—which regroups here *cultural, social* and *symbolic* forms of capital—instead of using *cultural* capital only.

Figure 1 (top diagram)

TOTAL CAPITAL +

BROADSHEETS

Investigative

General news

GOVERNMENT
BROADCASTERS
(Eg ABC, BBC)

COMMERCIAL TELEVISION

Politics

Sport

TABLOIDS

PRINT-FIRST

JOURNALISTIC
CAPITAL +

Business

ECONOMIC
CAPITAL+

World

SPECIALISED MAGAZINES

FREE NEWS WEBSITES

PAID ONLINE

Opinion/Editorial

PRINT-ONLY

Celebrity

Travel/Lifestyle

Motoring/Property/
Professional

Science/Health

WEB-ONLY NEWS SITES

WEB-FIRST

WEB-ONLY

COMMERCIAL RADIO

BLOGS

CAPITALS – TYPES OF MEDIA
ITALICS – *TYPES OF CONTENT*
Lower case – Reporting rounds

TOTAL CAPITAL –

Figure 2 (bottom diagram)

TOTAL CAPITAL +

EXCLUSIVE NEWS

Sports manager

Senior writer

PRINT AND WEB

Football

PRINT-FIRST

THE HINDU

*THE DAILY
TELEGRAPH*

Specialist
reporter

UNITED
KINGDOM

Cricket

THE *GUARDIAN*

INDIA

PRINT-ONLY

JOURNALISTIC
CAPITAL +

*THE TIMES
OF INDIA*

ECONOMIC
CAPITAL+

General
reporter

*THE SYDNEY
MORNING HERALD*

WEB-FIRST

AUSTRALIA

THE AUSTRALIAN

TWITTER

Online staff

Tennis

Subeditor

Rugby union

Rugby league

WEB-ONLY

WIRE CONTENT

TOTAL CAPITAL –

LARGE CAPITALS – Nations and organisations
SMALL CAPITALS UNDERLINED – Publishing and content types
Lowercase – Staff positions
Lowercase italics – Major sports

Figure 8: Field theory mapping of sports journalism. English makes use of Bourdieu's popular grid to analyse the field of sports journalism: in the first figure, he locates the subfield of sports journalism within the wider journalistic field, using a horizontal line that goes from the highest amount of journalistic capital (left) to the highest amount of economic capital (right), and a vertical one that spans from the largest amount of overall capital (top) to the smallest quantity available (bottom). This schema is reproduced in the second figure so as to map out actors and entities that can be found in the subfield of sports journalism in Australia, India and the United Kingdom. Source: English, 2016.

74

Additionally, a second research stream focuses on documenting the capacity of new entrants (e.g., bloggers, entrepreneurial journalists, news start-ups, technological companies) to either change or preserve the prevailing *doxa* in the field. This line of research is inspired by Bourdieu's belief (2005a, p. 39) that "to exist in a field (…) is to differentiate oneself". New entrants are then bound to either disrupt or abide by the prevailing norms or *doxa*, thus potentially reinforcing or changing the nature of both economic and cultural capitals. In that regard, Vos, Craft, and Ashley (2012) viewed political bloggers as new entrants in the journalistic field with a foothold, still, in the political field: as they engage in press criticism, they can be considered as playing a role in determining what constitutes cultural capital in the field, although the authors stress that, eventually, these bloggers seemed to have accepted the prevailing *doxa*. In a three-part study of the digital news start-up BuzzFeed, Tandoc, Jenkins and Foo evaluated that, as a new entrant, BuzzFeed was perceived by established players in the field as willing to maintain the prevailing *doxa* (Tandoc & Jenkins, 2017), although differences in terms of content (Tandoc, 2018), relationships with audiences and experimentation (Tandoc & Foo, 2018) were also to be found. Lastly, Usher (2017) brought forward that news start-ups do not fundamentally change the main *doxa* of the field, but rather challenges the *habitus* of journalists through the use of new technological products. For instance, news start-ups could consider that providing quality news to readers can be better achieved through the use of algorithms rather than by trusting a journalist's *gut feeling* (see Schultz, 2007), thus challenging their *habitus* of curating news items to inform the public.

3.3   Field theory for more interpretative and critical views

As detailed above, each of these frameworks brings forward concepts that can be useful when undertaking a practice-led study of automated news. ANT can be used at a descriptive level to give a "bottom-up" account of all the transformations that automated news is undergoing, thus pointing out to the overall direction that the "actor-network of automated journalism" is taking. However, ANT also comes with

potential limitations (see Ryfe, 2022) such as being too oblivious of any overarching social order. As Benson underlined (2017, p. 34), "its descriptions often lose sight of the (contextual) forest for the trees; (...) it fails to see patterned variation; (and) careful not to impose normative judgments, it refuses to draw obvious connections to real-world concerns and possible solutions". Furthermore, in light of all the challenges posed by datafication, Couldry argues, in his critique of the sole use of ANT (2020, p. 1146), that researchers and citizens alike should not lose track of the bigger picture and that, in this sense, "critical sociology must be revived". Field theory can help remediate these concerns in that it accounts for the influence that wider forces exert on both individuals and their immediate surroundings, but also bridges an important gap between structure and agency as the former is able to change the latter. As such, it will become my framework of reference to document more interpretative and critical views of automated news, which calls for further operationalisation.

3.3.1   Field theory's focus on tensions and rationale for using it with ANT

The attention given to structure and agency in Bourdieu's Field theory can also be found in other sociological concepts, most notably in Gidden's (1984) *duality of structures* model where institutions shape and are maintained by individuals' attitudes, but leave them enough room to manoeuvre so that they can in turn change them. In Bourdieu's work, though, the emphasis on tensions allows for a critical reading which is especially important in the context of datafication and artificial intelligence (see section 2.1.3): according to Anderson (2013, p. 1013), it brings "a vector of power dynamics to an area of socio-technical life (technological innovation) too often understood from within an 'all boats will rise' mentality".

    These forms of tensions are even more important to take into account as digital news is sometimes seen as a threat to news media's business models, and as new players like "net natives", bloggers and platforms are clearly challenging traditional news organisations (see Nielsen 2012, 2016; Bell *et al*., 2017; Rashidian *et al*., 2019, 2020; Vos, 2019; Nielsen & Ganter, 2022). Looking at the relationship between platforms and publishers for example, Facebook's changes to its news feed algorithm best exemplify those: in 2014, the social media giant announced a shift to

prioritising native videos on its news feeds, and in 2018, it revealed that interactions between friends would show up more often than Pages content, which included news publishers. As a result of the 2014 "pivot to video", it was shown that news organisations mostly complied with Facebook's change by posting more videos than before (Tandoc & Maitra, 2018), thus clearly demonstrating how yet another agent external to journalism that is driven by economic considerations (i.e., audience size and advertising revenues) is able to reinforce the *heteronomous* pole of the field. As for the 2018 adjustment—which at the time was pictured as being catastrophic for news publishers—some organisations said they were able to mitigate this while posting more content on Facebook, even if it appears that the sharp drop in engagement feared did not actually happen (Cornia *et al.*, 2018). The examples above show the type of structural pressures that social media platforms exert on journalism today, which add to all of the other tensions that already existed as a result of multimedia and newsroom convergence efforts (see Deuze, 2005; Singer, 2004).

There are, of course, limitations to using Bourdieu's Field theory in media research: one of them has to do with the lack of research that Bourdieu himself has performed in the domain of journalism, his main contribution (i.e, *On television*, 1996) being rather a short manifesto than an in-depth piece of sociological research (Neveu, 2005, 2007). Besides, there are also other shortcomings since Bourdieu can be perceived as being oblivious of the *autonomous* role played by the political field when it comes to subsidising the media (Benson, 2006), as it turns out to be the case with well-funded public service broadcasters or newspapers with low advertising revenues receiving financial help because they contribute to a healthy democratic debate. Finally, it can be advanced that media practitioners sometimes get so entangled in their own set of *autonomous* values—which can also be a professional elite's own "narrow opinions" that are not often exposed to self-criticisms—that it results in a disconnect between them and some legitimate democratic concerns (Schudson, 2005).

Because of their ontological differences (i.e., ANT's key tenet is to leave aside any preconceived ideas to follow the actors' trails while Bourdieu's Field theory encompasses a fair deal of determinism)—and also perhaps because of personal animosity between Bourdieu and Latour (see Kofman, 2022)—research combining

ANT with Field theory has so far been scarce: in a sense, this is being oblivious to the idea that some of the "sociology of the social" can work as "companion concept" to ANT as it is sometimes encountered at a later stage (Winthereik, 2020); moreover, Couldry (2016, p. 5) stressed the importance of seeing ANT as "one important item in the media theorist's toolkit that, like any tool, needs to be supplemented by others". A good illustration of ANT being used in combination with Field theory can be found in Prior's metaphor of a *prism* when looking at "glitch" electronic music, where he sets out (2008, p. 316) to "open the black box of technology as well as the well-regulated ballet of the field":

> To invoke the *metaphor of the prism*, sighting the actor network through the field and vice versa might give us some valuable insights into the strengths and weaknesses of both. This would be to create temporary intellectual adjunctions between two of the most advanced tools available, not necessarily to synthesize them, but to create points of friction in a spirit of mutual critical practice. In this process, we do not disavow the blind spots within contemporary theories, but rather deploy them in productive ways to reveal how the blanks might be filled.
>
> (Prior, 2008, p. 317, emphasis mine)

If ANT can be used at a descriptive level to document the changes that automated news is undergoing as it is embedded within a news organisation, Bourdieu's Field theory meanwhile requires further operationalisation as it focuses on more interpretative and critical dimensions, starting with a delineation of structure and agency. To do so, I will rely on Wu, Tandoc and Salmon's (2019b) Field theory interpretation of algorithmic automation in journalism, where they identify three areas to look at: first, outside the journalistic field, in the forms of influences that external structures such as political, economic, social and technological forces hold on the field; second, inside the journalistic field, in the types of cultural capital that journalists need to acquire in the automation age; third, in the reactions that journalists have in light of increased automation, which contributes to creating tensions within the field.

3.3.2 Operationalising Bourdieu's Field theory

In my review of scholarship written on journalism practice, I have stressed the importance of looking at whether computational journalists—or in this case teams of technologists and journalists working together on automated news projects—participate in changing or preserving prevailing journalistic norms, a consideration that echoes Bourdieu's focus on new entrants either disrupting or reinforcing the main *doxa* in the field, which hereby becomes an important axis in this study. As developed above, I will use Deuze's *ideal-typical values* (i.e., *public service, objectivity, autonomy, immediacy, ethics*) as a representation of this journalistic *doxa* as I believe they come the closest to articulating the specifics of these "unspoken rules" and are often used as a normative model in media studies (see section 3.2.1), then connect each of them to Wu, Tandoc and Salmon's own delimitation of structure and agency in their Field theory analysis of algorithmic automation in journalism (i.e., structures external to the journalistic field, accumulating cultural capital, adversarial reactions within the journalistic field). In doing so, I will come up with a set of reflections whose conclusions will be readily available in Table 3 so as to facilitate the identification of important themes to be addressed in this research[26]. It is important to stress that I understand cultural capital as an extended set of technological and journalistic capital, in line with Bourdieu's (2005b) and Schultz's (2007) comprehension.

    a.   Structures external to the journalistic field:

In automated journalism, one of the areas where the influence of external structures could impede on journalistic *objectivity* relates to the over-reliance on a unique data source to fill in a single template. For instance, although automated stories on sports or election results could directly feed on sports leagues' and governments' open data portals or APIs, it is nevertheless important to balance those with complementary data sources, so that the overall story remains impartial. This could be achieved through an assemblage of pre-written templates (see Caswell & Dörr, 2018), which

---

[26] These reflections were originally published in the same academic article I co-authored with my principal supervisor (Danzon-Chambaud & Cornia, 2021). There are reproduced here as-is or slightly altered.

could also feed on alternative databases, such as those of supporters' clubs or watchdog groups. To do so, journalists' *news habitus* in selecting balanced viewpoints and credible sources would be essential.

The over-reliance on a single data source could also hamper the *autonomy* of the journalistic field while making it more *heteronomous* to external influences like political or economic forces. Indeed, corporations and institutions with the means to maintain large data catalogues—and also that generated a lot of material that can be used as training data in machine learning models—are more likely to play an active role in the creation of automated news, as opposed to less affluent grassroot movements and citizen groups. As a result, these influential organisations could end up being overrepresented in automated news coverage and exert a form of control over journalists' capacity to "tell the stories they want".

To avoid being too dependent on external datasets, media practitioners could garner data using their own reporting skills. This could be done for instance through "structured journalism" (see Caswell & Dörr, 2018), which stands for the idea of reporting news in a data format, or to put it differently of turning "narratives into databases" (Anderson, 2018, p. 13). To uphold the value of *public service*, media practitioners could adopt this structured journalism approach to collect sensitive material: on the one hand, they could avail of their regular journalistic capital while filling Freedom of Information and Protection of Privacy (FOIP) requests to access valuable datasets and, on the other hand, engage with new forms of journalistic capital like organising crowdsourcing campaigns to collect a large amount of public interest data, as in *The Guardian*'s crowdsourced investigation on MPs expenses in 2009 (Guardian reporters & Guardian readers, 2009).

Another issue that relates to over-relying on external datasets has to do with the possibility that algorithmic errors could make their way into automated news. This is especially relevant with regard to *immediacy*, as demonstrated by the 2017 automated news imbroglio where the *Los Angeles Times* warned of an earthquake that actually occurred in 1925 (L.A. Times: L.A. Now, 2017): a revision of the exact location of a 1925 earthquake that occurred off the coast of California mistakenly triggered a United States Geological Survey alert that was sent to newsrooms across the country, thus prompting the *Los Angeles Times*' Quakebot to publish this information as such. If human verification is necessary to avoid such mistakes, new

forms of journalistic capital is meanwhile needed to verify stories that are published within minutes at scale, as in *Le Monde*'s 36,000 stories to cover the results of the 2015 regional elections in France or Tamedia's 40,000 news articles to report on the outcome of a 2018 Swiss referendum (Rédaction du Monde.fr, 2015; Plattner & Orel, 2019). This capital could translate for instance into computing skills to be able to program computational tasks that can involve advanced statistical calculations and text recognition mechanisms (e.g., optical character recognition), in order to avoid discrepancies in automated news reports.

Besides catching algorithmic errors, news workers also need to be aware of algorithmic biases that could creep into automated news stories. This is especially true of machine learning models, as computer software could potentially suggest pre-made sentences to include in a copy (see Lindén, 2017). At the same time, this could be an excellent opportunity to revamp the ideal of *ethics* while equipping media practitioners with a new *habitus* that would be made of ethical considerations on artificial intelligence and data science. These could be delivered either through the organisation's own standards and practices (e.g., the BBC's principles on responsible machine learning, the Bavarian broadcaster Bayerischer Rundfunk's AI Ethics Guidelines or the Norwegian media group Schibsted's FAST framework) or *via* a source that is authoritative enough in the domain of journalism (e.g., a potential addendum to the Society of Professional Journalists' Code of Ethics).

b. Accumulating cultural capital:

The adoption of automated news within newsrooms thereby prompts news workers to re-examine their journalistic capital, whether while re-emphasising their human potential by adding more context to the story or focusing on in-depth forms of reporting instead (van Dalen, 2012), or while engaging with a new form of "computational thinking", in other words solving problems in the newsroom while applying abstract reasoning similarly to computer programming (see Wing, 2008, Diakopoulos, 2011; Stavelin, 2013; Gynnild, 2014). Using computational thinking when programming automated news challenges *immediacy* notions as it calls for reconsidering what constitutes the *news habitus*: authoring templates for automated news requires predicting elements of the story in advance (see Thurman, Dörr &

Kunert 2017), a craft that is difficult to acquire as it necessitates to be familiar with abstraction, which could further entrench an effect of *hysteresis* among media practitioners.

At the same time, a *news habitus* that would involve predicting elements of the story in advance could impede on journalistic *objectivity*. For instance, news workers could feel cornered into choosing a "winning" and a "losing" side beforehand, which could further reinforce perceptions that the journalistic field is governed by *heteronomous* forces such as the political or economic fields, or caught into its own set of *autonomous* values that may sometimes collide with the "best practices of democratic government" (Schudson, 2005, p. 222).

Adopting a computational thinking mindset would certainly involve adding more programming skills to what constitutes the journalistic capital, which would strengthen the *autonomy* of "tech-savvy" journalists as they could program their own automated news as opposed to outsourcing it to an external NLG provider (which limits journalists' capabilities to have the upper hand on algorithmic models at use). Moreover, the *hysteresis* effect that could potentially happen between, on the one hand, "hacker journalists" (Usher, 2019) and, on the other hand, rank and file journalists with little or no programming experience could perhaps be mitigated through third-party self-editing tools (e.g., Arria NLG Studio, AX Semantics), which feature a "No-code" form of programming language.

Additionally, media practitioners equipped with an adequate understanding of computational thinking could adapt their existing *habitus*, which is made of an expert knowledge of journalism *ethics*, to address new ethical aspects that arise from the use of the technology, such as when "defamatory content slips through the cracks" (Lewis, Sanders & Carmody, 2019, p. 15). They could also ensure that journalistic standards and practices are well embedded into automated news scripts, and verify that they are adequately maintained and up-to-date. As raised above, a *news habitus* that includes a good understanding of artificial intelligence and data science ethics will help them in this task.

Finally, adding programming skills to journalistic capital could open new avenues in the area of *public service* journalism, most notably with regard to "algorithmic accountability reporting", a new type of journalistic investigation that looks into black box algorithms in order to reveal "the power structures, biases, and

influences that computational artefacts exercise in society" (Diakopoulos, 2015, p. 398). As algorithmic accountability reporting is based on the use of reverse engineering—a set of techniques that serves to investigate the input–output structure of algorithms—journalists equipped with programming skills to build their own automated news could then engage with this new investigative format, leading them to reveal, among others, potential biases in hiring and credit scores algorithms or in predictive policing software, thus consolidating the *autonomous* pole of journalism.

c. Adversarial reactions within the journalistic field:

Looking at potential struggles within the journalistic field, a first type of tensions that could arise from the implementation of automated news within newsrooms relates to *autonomy*. Indeed, as news work involves editorial staff, but also technologists and businesspeople (see Lewis & Westlund, 2015b), the confrontation between each of their *habitus* could result in increased tensions within media organisations. For instance, although media practitioners and technologists share similarities in their respective *doxa* (see Wu, Tandoc & Salmon, 2019c), their views could be conflictual due to misunderstandings around what constitutes the boundaries of journalism (see Lewis & Usher, 2016). Moreover, as the economic appeal of automated news could translate into less human employment (see Kim & Kim, 2017), this could potentially fuel dissensions between editorial staff and businesspeople.

Such a reduction in human employment could also hamper less profitable forms of coverage, such as local news, and ultimately impede on the development of *public service* journalism. For instance, from a media management point of view, the savings occasioned by the launch of automated news products—coupled with the possibility to tailor it to niche and geo-specific audiences (see Lokot & Diakopoulos, 2016)—could in fact constitute an strong economic incentive that would eventually threaten the livelihood of specialised journalists and local correspondents, thus creating an effect of *hysteresis* among those unable to adapt.

Automated journalism is therefore likely to introduce a considerable reshuffle within newsrooms. A positive outlook would be that automated news relieves media practitioners from the challenges associated with *immediacy*, executing rote tasks for them like writing repetitive news recaps so that they can focus on more demanding

forms of journalism (van Dalen, 2012; Clerwall, 2014), what would ultimately strengthen the *autonomous* pole of the field. These more demanding tasks could include traditional formats like investigative or international reporting, in-depth forms of news reporting where journalists participate in the story (e.g., narrative, immersion and "Gonzo" types of journalism) and new formats tackling growing datafication in society, such as advanced data journalism or algorithmic accountability reporting. That said, if reporters are instead assigned to stories that reflect the personal views of media owners or the priorities of the marketing and advertising department (e.g., "clickbait" stories or native advertising), this would reinforce the *heteronomous* pole instead, and trigger an effect of *illusio* among news workers.

Another area of struggle that relates to the use of automated news deals with *objectivity* as new forms of co-authorship gradually emerge, either through media practitioners reworking NLG-powered first drafts (Wölker & Powell, 2021) or potentially through pre-made sentences generated with machine learning that could be dragged and dropped into a copy (see Lindén, 2017a). It is then important that media practitioners remain in control of the story, using their *news habitus* of critical thinking to make sure that the story is overall objective—especially when it draws on a single data source—and that no algorithmic biases creep into the final automated copy, thereby making the field more *heteronomous*.

To conclude, as previously mentioned, the introduction of automated journalism within newsrooms brings new perspectives on journalism *ethics*, thus fostering discussions on a potential renewal of journalistic standards and practices. If this conversation takes place within news organisations, journalism research centres and professional associations, the *autonomous* pole of the field would be reinforced, but if it is too closely tied to external organisations—especially like Big Tech companies (see Haapanen, 2020)—there is then an increased risk that *heteronomous* forces penetrate the field of journalism.

These reflections illustrate the many ways through which automated journalism could change and reinforce the prevailing journalistic *doxa*. In summary, the influence external forces exert on the field could be visible in an over-reliance on external datasets, the accumulation of cultural capital would have to do with media

practitioners acquiring a computational thinking mindset and adversarial reactions within the field could relate to potential tensions between editorial staff and, one the one hand, technologists and businesspeople in news and, on the other hand, players external to the field, in this case Big Tech companies. In my research thesis, I will use these considerations as a general thread that will guide my more critical and interpretative chapter; they are also summarised in Table 3 so as to have important themes readily available when I address this part of my empirical work.

That said, it is important to stress that these considerations should also be weighed against existing journalistic practices that present similar issues. For instance, relying on one data source is essentially similar to relying on a limited number of authoritative and official sources for routine news, which ultimately hampers diversity in media coverage (see Gans 1979; Molotch & Lester 1974). Likewise, trying to predict elements of the story in advance when authoring templates—and risking to pick a "losing" and a "winning" side beforehand—corresponds in a way to "prep copies" that journalists write ahead of time, which could be problematic should reporters introduce perspectives into their copies: this is indeed a difficult style of writing to acquire, as shown in the work of journalists writing "advance obits" (Adams, 2022).

Table 3. Important themes to consider when investigating automated news deployment.

	Public service	Objectivity	Immediacy	Autonomy	Ethics
Structures External to the Journalistic Field	• Are media practitioners able to employ their traditional journalistic skills to collect public interest data? • Are they able to explore other ways to get public information data?	• Is a diverse set of data sources used in the conception of automated news? • Is an assemblage of templates being used or just a single one? • Are alternative data sources taken into consideration?	• What is the extent of human verification in the production of automated news? • Are computational tasks employed to verify content when automated news are published on a large scale?	• Are organisations able to build and manage large open data catalogues overrepresented in automated news reports? • In advanced machine learning models, is it also the case for organisations that generated a lot of training data? • Are less affluent groups or movements not covered on these grounds?	• Are media practitioners well aware of algorithmic biases? • Are they trained in ethics of artificial intelligence and data science?
Accumulating Cultural Capital	• Are journalists with the computing background to program automated news also able to engage with algorithmic accountability reporting?	• When authoring templates, do media practitioners refrain from picking a "winning" side and a "losing" side beforehand?	• How do media practitioners handle predicting elements of the story in advance when authoring templates? • How familiar are they with the concept of abstraction at the core of computational thinking?	• Are journalists with a computing background able to program their own automated news? • Are media practitioners with little or none programming background able to use third-party NLG tools or "No-code" development platforms?	• Are media practitioners able to adapt their expert knowledge of core ethical values in journalism to new ethical challenges that arise from the utilisation of automated news? • Do media practitioners supervise the incorporation of journalistic standards and practices into automated news scripts? • Do they also make sure that those are well maintained and up-to-date?
Adversarial Reactions within the Journalistic Field	• Does the deployment of automated news threaten the sustainability of less profitable areas of journalism? • Does it impact the livelihood of journalists specialised in these areas?	• When complementing an automated draft, do media practitioners remain in control of the story while exerting their critical thinking? • In advanced machine learning models, do media practitioners prevent algorithmic biases from creeping into the final copy?	• Are media practitioners able to engage with formats that demand more time as automated news take on routine tasks? • Are they redirected to stories that reflect the priorities of media owners or of the marketing and advertising departments?	• How do media practitioners views on automated journalism collide with those of technologists and business people? • Is the displacement of human journalists likely to create tensions?	• Does the discussion around revamped ethical values happen within professional journalistic forums, or is it too closely tied to external organisations?

4   METHODOLOGY

In this part, I will delve into the specifics of my methodological choices for this research thesis. I will first explain my ontological choice of taking a critical realist approach, then detail my rationale for using an iterative-inductive and case study design. I will also specify why I am using semi-structured interviews—which are complemented by elements of a netnography—as well as thematic analysis, before giving a detailed account of the procedures that I have chosen to follow.

4.1   ONTOLOGICAL CONSIDERATIONS AND RESEARCH DESIGN

To start, my research work takes a critical realist ontological approach, in that the epistemological traditions that I follow are related to both positivism/objectivism and constructivism/relativism. Critical realism reunites elements from both of these schools while remaining an independent position in itself and not just a mere combination of the two (Clark, 2008; Kozhevnikov & Vincent, 2019). According to critical realism, positivism is viewed as failing to consider the social nature of knowledge, while constructivism can be thought of as putting too much emphasis on it (Clark, 2008). Critical realism therefore comes as a third avenue that offers an ontology that follows neither of these two approaches (Elger, 2010).

The work of Bhaskar (2008) can be seen as instrumental in establishing a critical realist position as he made the distinction between an *intransitive* and a *transitive* dimension in the observation of knowledge: the *intransitive* dimension corresponds to knowledge that objectively exists and can be observed as such (e.g., beings, relations, processes, etc.) while the *transitive* one deals with sociocultural concepts that can be only be captured through human interpretation (Benton, 2004; Houston, 2014). Bhaskar also divides reality between an *empirical*, an *actual* and a *real* or *causal* level: the *empirical* is concerned with observations we can directly sense, experience, witness and perceive, for instance listening to a classical concert or witnessing that there are fewer women in engineering schools; the *actual* accounts for events, processes and outcomes that are beyond our direct scope, like a concert

happening in a faraway location or an invisible problem of institutional sexism in engineering that may explain why fewer women are attending this curriculum; and the *real* or *causal* addresses underlying powers, structures or mechanisms that contribute to triggering the *actual* and *empirical* levels, for instance stereotypes and gender roles that may limit women's aspirations to study engineering (Clark, 2008; Elger, 2010; Houston, 2014; Kozhevnikov & Vincent, 2019). It is important to note that, as the *actual* and *real* levels both build on human interpretation (i.e., *transitive* dimension), they are at greater risk of being fallible (Benton, 2004; Clark, 2008).

My research takes a critical realism position in that it will look into both the *intransitive* and *transitive* dimension: the epistemological tradition that is being followed by Bourdieu can be seen as belonging to the *transitive* dimension as it addresses *real* perceptions (e.g., hidden social mechanisms) while the "follow the actors" motto of ANT makes it more an *empirical* approach that sides with the *intransitive* dimension. I will also document the *actual* level while focusing on occurrences that are not directly visible to me, like computer software or organisational polices. For the sake of simplicity, though, I should refer to the *empirical*, *actual* and *real* levels as micro, mezzo and macro perspectives in the rest of this study.

In terms of research design, my thesis breaks away from both deductive design, which usually makes use of theory to test an hypothesis, and inductive design, which aims at formulating new theoretical propositions with as much of an open mindset as possible, in that it reflects a *sophisticated inductivism* or *iterative-inductive* type of design (O'Reilly, 2009, 2011): in iterative-inductive design, theory is used throughout the whole research project, allowing the researcher to move back and forth between research phases rather than progressing in a strictly linear manner. According to O'Reilly (2009), iterative-inductive design is based on the idea that theory can be viewed as a "pre-cursor, medium and outcome of ethnographic study and writing" (Willis & Trondman, 2000, p. 7), a concept that now applies to other types of research so as to facilitate (Ezzy, 2002, p. 10) "an ongoing simultaneous process of deduction and induction, of theory building, testing and rebuilding".

Theory is all the more important to use in that it allows for *analytic* or *theoretical* generalisation, a way of achieving external validity other than through

statistical or *numerical* generalisation. Theoretical generalisation can be used to evaluate the findings of a study against previously developed theories, so as to support, contest, elaborate or refine their relevance (Schwandt, 2007; Akremi, 2020). As pointed out by Yin (1994), an appropriate structure that suits theoretical generalisation is case studies, which can be described as in-depth or "thick" descriptions of one or several instances of a phenomenon (Blatter, 2008). Case studies can either be *single* case studies or *multiple/cross-case* studies (Yin, 1994; Gerring, 2007): they can come under a *holistic* form, whereby a case is being investigated in its full entirety, or under an *embedded* form, where the object of study is divided into units or subunits. That said, Yin stresses that single and multiple case studies differ in their rationale for case selection: single case studies can be selected based on whether they represent a *critical* case that best corresponds to a theory that is being tested; an *extreme* or *unique* case that depicts a situation particularly rare or that is unfolding under extreme circumstances; or a *revelatory* case where the researcher has access to a site or a phenomenon that was not accessible to the scientific community before, thus leading to exclusive insights (Bleijenbergh, 2010; Xiao, 2010). As for the rationale for selecting elements of a multiple case study, this is linked to the development of a rich theoretical framework, so that findings can be generalised to new cases (Yin, 1994). On a last note, Gerring suggested that single and multiple/cross-case studies could be conducted in a complementary way, even advising (2007, p. 12) that "researchers may do both and, arguably, must engage both styles of evidence".

4.2   Semi-structured interviews and thematic analysis

To investigate the considerations I set out in this chapter, interviews with media practitioners, executives and technologists appear to be the most pertinent research method: on the one hand, conducting those *via* computer-mediated means (see Hansen & Machin, 2013) was the only way of completing this research project because of COVID-19—which can be a rationale on its own (Mason, 2002)—and, on the other hand, interviews are an appropriate tool that can be wielded to collect both informative and interpretative material: in line with investigating both the *intransitive* and *transitive* dimensions, they can then either be used (Brinkmann and

Kvale, 2018) as a "miner", where the interviewer is concerned with gathering factual accounts only (i.e., micro level), or as a "traveller", which implies working with more interpretative views (i.e., mezzo and macro levels). In my research thesis, I will be using interviews both as a "miner" and a as a "traveller", essentially to reflect the critical realism position that I am taking.

To strike a balance between collecting factual evidence and personal impressions, semi-structured interviews—sometimes called in-depth interviews (Cook, 2008)—can be seen as an appropriate format, as opposed to unstructured interviews, which are better for documenting personal experiences only (e.g., oral histories), or structured ones, which more adapted to just gathering objective information (e.g., survey interviews). Aside from touching on a list of key issues and themes that are mentioned in a questionnaire or *aide-mémoire*, semi-structured interviews also enable the researcher to digress from those lines so as to obtain other types of insights (Leonard, 2003). Not only semi-structured interviews allow for a more natural tone of conversation to take place (Boyle & Schmierbach, 2019), they are also a good fit in situations where concepts and relationships are well understood between the interviewer and the interviewee (Ayres, 2008a). Therefore, using semi-structured interviews with a professional like a journalist can help access information that is not necessarily obvious in the first place, and cannot easily be shared in other formats (e.g., focus groups), for instance talking about professional concerns (Hansen & Machin, 2013).

That being said, interviews also come with their own set of issues, most notably when interviewees answer questions the way the researcher wants to hear it (Berger, 2000), making it comparable to the *Hawthorne effect* where research participants change their behaviour according to whether interest is being shown in them or when they feel they are being observed (James & Vo, 2010). Although it is not possible to entirely take away potential biases that may occur over the conduct of the interview—mostly because of social interactions—semi-structured interviews can nevertheless help mitigate those by way of standardising questions in the *aide-mémoire* (Mason, 2002). To be able to carry out this type of interview, I undertook training while attending a methodology workshop at Dublin City University's Qualitative Research Summer School.

In complement to using semi-structured interviews to access qualitative insights, interpreting them through thematic analysis can also be key, in that it helps account, again, for the same critical realist approach. According to Boyatzis (1998), thematic analysis can be understood as a process of sorting qualitative data through coding procedures that are centred around themes that are either *manifest*, meaning that they describe the patterns found in the data as such, or *latent*, which implies trying to interpret what is behind those results. By the same token, Braun and Clarke (2006) envisioned thematic analysis as being either *realist/essentialist* when describing reality as such, *constructionist* when examining it through the lenses of discourses held in society, or *contextualist* when it sits somewhere between the two, making it closer to the critical realism tradition. In this PhD dissertation, I will follow the latter as the data that I am collecting reveals both observable *manifest* patterns around automated news usage and more *latent* dimensions that have to do with media practitioners' own interpretations.

Following an *iterative-inductive* type of design, themes used in thematic analysis can also match a theoretical construct that was previously outlined (Boyatzis, 1998; Morey Hawkins, 2017) and as such can be visible in the research questions and questionnaires (Lapadat, 2010). This is especially true of semi-structured interviews, where themes can somehow be anticipated in questionnaires (Ayres, 2008b). However, Braun and Clarke (2006) cautioned against using themes this way, as it runs the risk of discarding the in-depth analytical work that is needed to find out about those in the first place.

4.3   ACCOUNT OF PROCEDURES

In accordance with Gerring's recommendation, this research thesis is based on a combination of single and multiple case studies. While the two single case studies came as a result of opportunities that were presented to me as a researcher, the multiple case study reflects the need for a wide-reaching study of automated news that was mentioned in the Introduction. The single case studies were selected on the basis that they were illustrative of a *unique* and a *revelatory* use of automated news: the *unique* case study illustrated how a selected number of media organisations resorted to automated news to cover COVID-19, a rare event during which the

technology was used under unprecedented circumstances—which I turned into an industry report for the Tow Center for Digital Journalism at Columbia University (see Danzon-Chambaud, 2021b) where I was a Knight News Innovation Fellow—while the *revelatory* case shed light on a range of experiments that took place at the BBC— which were summarised into a white paper (see Danzon-Chambaud, 2021c) I wrote as part of a secondment with BBC R&D[27]. Even though this secondment was done virtually because of COVID-19, I enjoyed access to some of the BBC's internal systems, participated in a few training and socialising events and, more importantly, was put in touch with interviewees for my research project. For the COVID-19 case study, I used a convenient sampling strategy as I reached out to the few news organisations I could see were implementing automated news at the beginning of the pandemic. Initially, I talked to nine organisations for my Tow industry report, but then found out—while conducting interviews for my multiple case study—that three others were also using automated news to cover COVID-19: in this PhD dissertation, they all figure together in chapter 5.1. As for my BBC case study, I rather used a snowball type of sampling as I asked my interviewees for potential additional respondents to this project. Although it is evident that my BBC case study is a single one because it focuses on one media organisation's use of automated news and can be considered *holistic* because it tackles all of BBC's experiments of this type at once, it is important to stress that this is equally true of my COVID-19 case study, which—despite coming under an *embedded* form—covers one single event and is therefore theme-centric.

Aside from these two single case studies, I also resorted to a multiple case study where I analysed the use of automated news across groups of countries and media types. To do this, I have chosen a sampling strategy that reflects Hallin and Mancini's (2004) media system typology so as to be able to ground this case study on their theoretical understanding of *differentiation* and *de-differentiation* in the news industry, which also echoes Bourdieu's reflections on commercial homogeneity within the news media (see section 3.2.2). Moreover, Hallin and Mancini's typology has often been used as a guiding framework to draw strategic samples of news organisations that spread across groups countries (see Cornia, Sehl & Nielsen, 2019,

---

[27] My Tow Center (i.e., COVID-19 case study) and BBC findings are reproduced as such or under a slightly modified form in this PhD dissertation.

2020; Menke *et al.*, 2018; Sehl, Cornia, Graves & Nielsen, 2019; Sehl & Cornia, 2021; Sehl, Cornia & Nielsen; 2021; Van den Bulck & Moe, 2018).

As for choosing media types, I have decided to include news agencies, newspapers and public service media for the following reasons: first, the use of automated news at news agencies is well established, especially in Europe (see Fanta, 2017); second, it can be argued that newspapers are more likely to engage with this form of technology, as their business model that is under threat because of the digital turn forces them to be more innovative, as opposed for instance to commercial broadcasters that can still rely on stable advertising revenues and on other types of incomes (Cornia, Sehl & Nielsen, 2019); third, public service media can be considered leaders in providing "thorough" data journalism pieces to audiences (see Borges-Rey, 2016; De Maeyer *et al.*, 2015), especially as data journalism experts are more likely to be hired at public service broadcasters in Germany (Beiler, Irmer & Breda, 2020) and as public service media in Australia developed their own in-house solutions (de-Lima-Santos, Schapals & Bruns, 2021): this can let us posit that the kind of programming skills that is at use in data scraping activities can also be leveraged to set up automated news. For this case study, I have relied on purposive sampling to select 18 news organisations, with each pair representing a different combination of media types and media systems (see Table 4). This sample includes news organisations that were also featured in my COVID-19 case study, but this time I am touching on their other uses of automated news as well. In addition, my BBC findings are also included here. The resulting *embedded* format is representative of organisations I deemed to be characteristic enough of their media system to be included in this sample, but also—of course—that were available and willing to take part in this research study.

Table 4. News organisations studied based on media systems and media types.[28]

---

[28] There are a few shortcomings in this selection: first, Australia is not examined in Hallin and Mancini's work for practical reasons, but they do specify that both Australia and New Zealand have close connections to Western European countries; second, Hallin and Mancini describe Belgium as a mixed case that sits between the Democratic Corporatist (i.e., North/Central) and the Polarised Pluralist (i.e., Mediterranean) models: in this study, I have included it under the latter category as Rossel and its subsidiary Sudpresse own titles in France and in French-speaking Belgium; third, there is the issue of limiting myself to Western news organisations, as discussed in my literature review and also in subsequent work on media systems (see Hallin & Mancini, 2011); finally, even if the Spanish newspaper *El Confidential* does not have a print edition, it defines itself as a "digital newspaper" (*diario digital*), which is why it is included here.

Media systems	News agencies	Newspapers	Public broadcasters
North Atlantic	Associated Press (United States)	Washington Post (United States)	BBC (United Kingdom)
			ABC (Australia)
	Reuters (United Kingdom)	The Times (United Kingdom)	
North/Central	STT (Finland)	Stuttgarter Zeitung (Germany)	YLE (Finland)
	NTB (Norway)		Bayerische Rundfunk (Germany)
		Tamedia (Switzerland)	
Mediterranean	AFP (France)	El Confidential (Spain)	France Bleu (France)
	ANSA (Italy)	Rossel/Sudpresse (Belgium/France)	RTVE (Spain)

In total, 23 news organisations figure in this research project, which is made of 30 interviews (average length: 00:35:30) with editorial staff like journalists, editors and managers; executives like directors and C-level managers (who are representative of the business side, at a managerial level); and technologists like software engineers (see Appendix B for my interviewees' news organisations, roles and genders, and their interview dates and durations). Among them were 8 BBC staffers that I could gain access to thanks to my secondment. Interviewees were contacted by email or *via* social media, a gatekeeper's approval being sometimes needed[29]. All in all, I found that my respondents were relatively easy to get to, my position as a Tow Fellow and my BBC internal email for the time of my secondment being certainly of help. These interviews were conducted between June 2020 and April 2021, with one of my BBC interviewees also performing a virtual walkthrough of the self-editing tool they used (i.e., Arria Studio) on April 9th, 2021. To determine questions to be asked, I relied on key considerations summarised in Table 3 in order to select those that were most appropriate to the news organisation or the professional profile of the person I was interviewing, and then adapted them into individualised questionnaires that also included more factual questions. For example, the key consideration *How do media*

---

[29] These "gatekeepers" could be, for instance, members of the marketing and advertising team, editorial managers or media executives.

practitioners handle predicting elements of the story in advance when authoring templates? was adapted, in my questionnaire for the *Washington Post* (see Appendix C), into the following: "When writing templates, how do they (media practitioners) balance predicting elements of the story in advance with the uncertainty of news?"

My exchanges were rather smooth[30], my interviewees generally knowing what I was asking about and not being caught off-guard (they were given an indication of what will be discussed, but were not handed the interview questions in advance). Only one time I had to redesign my questionnaire on the spot when I learned the organisation I was interviewing used machine learning instead of the template-based approach that I had prepared myself for, but this did not adversely affect the conduct of the interview. I also asked for extra clarifications in follow-up emails when needed. As this study is exploratory by its nature, semi-structured interviews helped me get to the kind of rich qualitative material I was looking for, unlike for instance an online survey that would involve too much set answers or focus groups where participants would not feel as comfortable to share their personal experiences (see Hansen & Machin, 2013). To complement these interviews, I also analysed material published online (e.g., blog posts, trade publications, etc.) so as to have a better overview of the way automated journalism is implemented: these are mentioned as such in chapters 5 and 6; otherwise, information comes from statements collected over the course of my interviews. In addition, my more descriptive empirical part features screenshots of automated news software or material that was found online or forwarded to me by research participants. All in all, the assemblage of all these data that were collected through online interviews, virtual walkthrough, analysis of online content and screenshots—but also email conversations—can be considered as as many *sources of evidence* as identified by Yin (1994): these are documents, archival records, interviews, direct observations, participant observations and physical artifacts, which bears resemblance to the material at hand. In the end, because some of these elements exclusively dealt with online collection, it is then appropriate to speak of a *netnography*, which is understood (Kozinets, 2016) as a "specific approach to conducting ethnographic research that uses the archival and communications

---

[30] Interviews with French speakers were conducted in French, as it is my mother tongue. I also took care of translating it into English.

functions of contemporary Internet-based technologies such as mobile phones, tablets, and laptop computers" and can be made of textual, graphic, audio, photographic and audio-visual elements.

I then tapped into this rich qualitative material using thematic analysis, resorting for this to the three themes uncovered in the first half of this chapter (i.e., over-reliance on external datasets, need for media practitioners to acquire a computational thinking mindset, conflicts within and outside the journalistic field). Using NVivo qualitative data analysis software, I regrouped some of my interviewees' statements around three nodes that each represented one of these themes. In a second step, I reflected on the material at hand and configured more specific sub-nodes based on patterns I could identify in my interview data (e.g., dealing with multiple levels of government for relying on external datasets; developing an abstract thought process for acquiring a computational thinking mindset; opposition between technologists' and journalists' own workflows for conflicts within and outside the field), then reassigned some of my interviewees' statements accordingly. As mentioned earlier, these themes and sub-themes were then used as a thread in chapter 6, which constitutes my more interpretative chapter. Finally, my findings were summarised and analysed in my conclusive chapter, using elements of Field theory and ANT for this, so as to best answer the research questions I have set out earlier.

On a last note, my submission to Dublin City University's Research Ethics Committee stressed that, as media professionals, research participants should be aware of the sensitivity of the material they provide and would not usually provide researchers with information they do not want to be published. Their names were not divulged so that they could speak more freely, although it is most likely that their hierarchy knew that they were participating in this research project. To avoid putting them in an uncomfortable situation, interviewees were given the opportunity to review some of their statements that dealt with potentially sensitive or unclear information, but not my own interpretation over what they shared. Overall, it can be said that they view automated news in a much positive light, even if they did not hesitate to share failures as well. This makes the use of theory even more relevant, as it provides the necessary lenses to have a critical outlook that can help understand aspects that were

not obvious in the first place. Finally, sex and gender were not considered to be particularly relevant in this study: as such, interviewees were not asked to disclose their gender, but in a strictly binary sense it turned out that 25 of them were men and 8 were women, thus reflecting a gender gap that could be further investigated.

As mentioned earlier, conducting remote semi-structured interviews was the only way to see this research through because of COVID-19. For a time, direct observations were envisaged through newsroom ethnography, which gives the researcher access to the cultural and naturalistic setting that surrounds media production and can be used to document the routines and everyday thoughts of news workers (Berger, 2000; DeWalt & DeWalt, 2011; Hansen & Machin, 2013). However, as the pandemic appeared to last for longer than expected in the middle of 2020, it became evident that this research method would not be available to me over the course of my PhD. This implied limitations in that I have not been able to hear bits of conversation and observe participants' behaviours so as to have a more genuine appreciation of media practitioners' views on automated news, in contrast to the more conventional tone used during interviews. As a result, somewhat confrontational statements that could testify, for instance, of potential tensions within the newsroom were much harder to obtain, which is reflected in section 6.3 being shorter than the other ones. This also meant that I could not speak with all of the teams I wanted to talk to: this is why some of the business branches (i.e., marketing and advertising department) are missing in this study, as their representatives were not identified as key persons to talk about automated news by their organisations.

5   COVID-19, BBC AND CROSS-NATIONAL CASE STUDIES

In this chapter, I will start my empirical analysis with a descriptive account of how automated news is being used in the three case studies that I am investigating: first, during COVID-19; second, as part of series of BBC's experiments; third, across news organisations that are based in different countries. Due to the lower number of organisations studied—or sole focus on one of them—in my two single case studies (i.e., COVID-19 and BBC), I will analyse usages based on patterns that I could directly observe (i.e., different ways of using automated news during COVID-19, minor and major experiments at the BBC). That being said, in my cross-national multiple case study, which features a higher number of organisations, I will resort to ANT so as to tell apart automated news being used as *intermediaries* (i.e., initial intent is maintained and it essentially does what it is supposed to do) from automated news being used as *mediators* (i.e., something *more* is added to existing practices). As this part is essentially about describing the utilisation of automated news in these three case studies, netnographical elements (i.e., screenshots and material published online) will be extensively featured here—along with interviews—so as to give the best overview possible.

5.1   USING AUTOMATED NEWS TO COVER COVID-19

In a sense, COVID-19 could be the perfect story to automate. When the virus spread globally at the beginning of 2020, governments and health authorities made accessible a considerable amount of open source data, generally available through structured datasets or APIs. These statistics contained critical information like the number of deaths and patients in intensive care units as well as 7-day incidence rates. The availability of these structured data, some of which could fit into predictable or templated story frames (see section 1.2.2), enabled some media organisations to use automated news in their coverage of the pandemic. This section provides examples from 12 outlets from ten countries (see Table 5), including 10 that were already using automated news for other types of coverage (i.e., ANSA, AFP, Bayerischer Rundfunk,

Bloomberg, Canadian Press, *Helsingin Sanomat*, NTB, RADAR, Reuters and Tamedia) and two that experimented with the technology for the first time (i.e, Omni and *The Times*). For the most part, these organisations developed their own automated news products in-house, among which figures a machine learning system designed by Bloomberg's engineers. In two cases (i.e., Tamedia and RADAR), they used a third-party self-editing tool so that journalists can design their own stories, and in another two (i.e., Omni and ANSA) the production of automated news was outsourced to an external content provider. As developed below, the use of automated news to cover COVID-19 essentially helped fulfil one of two goals: either to provide readers and media clients with user-facing interfaces rounding up the latest numbers on the virus, or to rearrange or come up with new types of workflows focused around setting it up or working directly with it.

Table 5. News organisations examined* in COVID-19 case study.

Country	News organisations
Canada	Canadian Press
Finland	Helsingin Sanomat
France	AFP
Germany	Bayerischer Rundfunk
Italy	ANSA (Applied XL)
Norway	NTB
Sweden	Omni (United Robots)
Switzerland	Tamedia (*Wordsmith*)
United Kingdom	RADAR (*Arria NLG Studio*), Reuters, The Times
United States	Bloomberg

*External content providers and third-party self-editing tools (in italic) are mentioned in parenthesis.

5.1.1 Quantifying the pandemic for readers and media clients

The first use of automated news to cover the COVID-19 pandemic has to do with providing a statistical overview of the spread of the virus to media clients and

audiences alike, either through user-facing interfaces such as dashboards and newsletters or through an extended range of news products. For instance, ANSA, Reuters, Tamedia and *The Times* (see Figure 9) all featured local, national and international dashboards that summarised the latest COVID-19 statistics with automated text and visualisations, like area charts detailing the spread of the pandemic in Italy (i.e., ANSA), bar charts showing a worldview of vaccination progress (i.e., Reuters and Tamedia) or a map illustrating weekly cases numbers in the United Kingdom (i.e., *The Times*). Whereas *The Times* has used automated graphics in the past, using automated text generation was a first: "It was really the first time where we felt the need for such a long-term investment in something that was automated", explained a computational journalist at the newspaper.

Figure 9: *The Times*' **COVID-19 dashboard.** Screenshot of the COVID-19 tracker page at *The Times*. It features automated text and visualisations that give a breakdown of the latest coronavirus statistics in the United Kingdom. Source: *The Times*.

This type of interactive dashboard also took the shape of a newsletter, such as the one launched by the Bavarian broadcaster Bayerischer Rundfunk, which features automated text as well as tables, maps, and charts. This COVID-19 newsletter (see Figure 10) draws on raw data released by the Robert Koch Institute and is updated on a daily basis. It shows the spread of the virus in Bavaria and Germany and informs readers on a selected range of indicators such as total cases, deaths and recoveries as well as 7-day incidence rates at local levels. Similarly, automated stories generated daily at the Canadian Press were used to feed an automated newsletter that would wrap up the most recent numbers for media clients. It included pandemic-related data like the number of confirmed cases, recoveries and deaths, but also economic data drawn from other automated coverage at the news agency in order to account for the financial repercussions of the virus. Automated charts were also created to keep track of new cases numbers all across Canada. The same could be observed at AFP, where an automated COVID-19 roundup featuring a world overview was produced on a daily basis for media clients and came along with automated visualisations, some of which included short animated videos showing the evolution of the virus (see Figure 11).

Another way of providing a statistical overview was through a new range of products that were available to media clients. For instance, the Norwegian news agency NTB delivered automated news on COVID-19 directly through its wire service (see Figure 12), but also through an API that news organisations could use to set up their own interactive products like live blogging platforms (smaller clients could also use NTB's own platform for this). These automated pieces included health-related statistics such as the status of vaccinations across Norway, as well as unemployment and furlough figures caused by the pandemic. The news agency also provided automated visualisations that were self-updating.

Figure 10: Bayerischer Rundfunk's COVID-19 newsletter. Screenshot of the automated COVID-19 newsletter set up by the Bavarian broadcaster Bayerischer Rundfunk. It informs readers on the spread of the pandemic both in Bavaria and Germany. Source: Bayerischer Rundfunk.
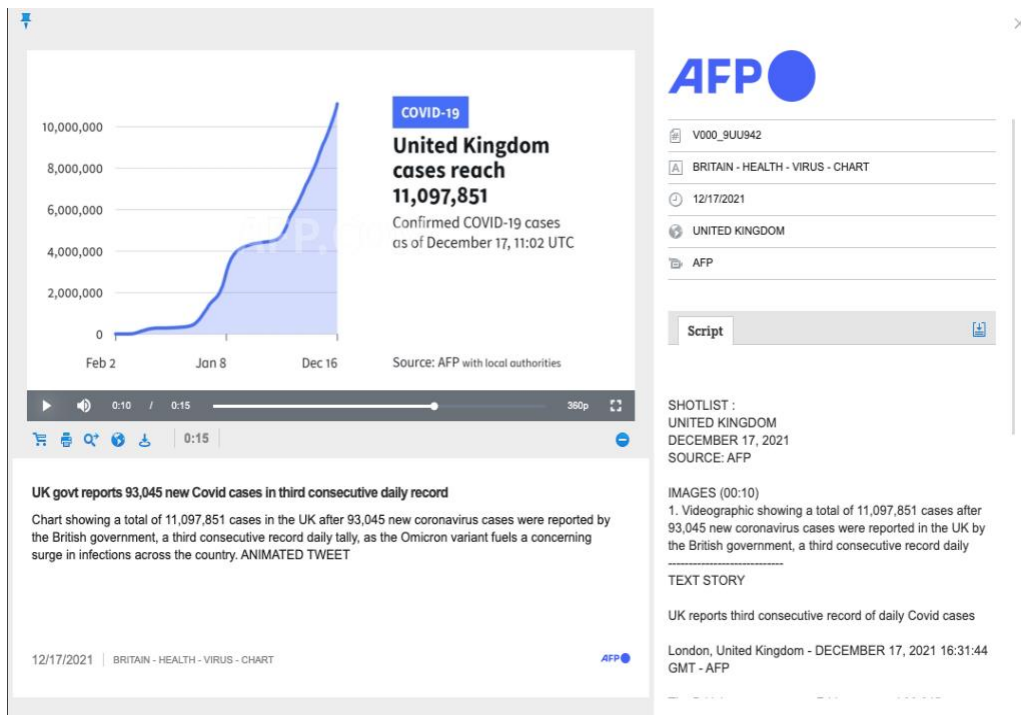
Figure 11: AFP's automated video chart. At AFP, automated visualisations could take the shape of short animated video charts that were used to document the spread of COVID-19 in a given area, like in the United Kingdom in this case. Source: AFP.
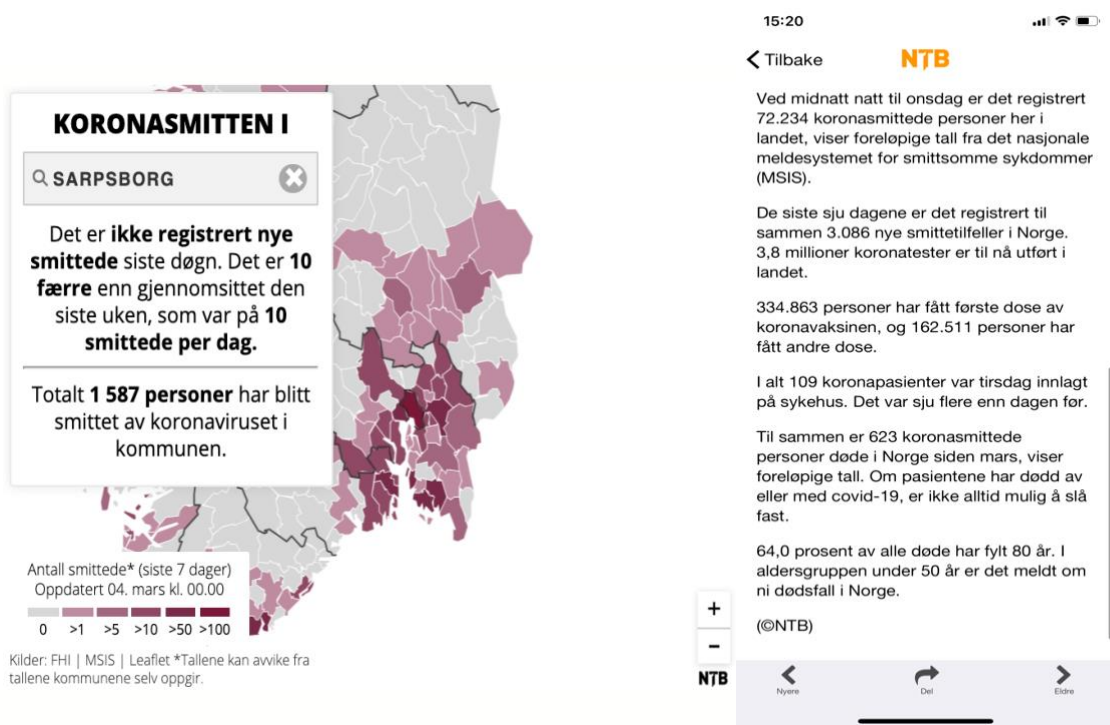
Figure 12: NTB's automated content. An example of automated map and text on COVID-19 that NTB generated. NTB's automated stories included health-related statistics, but also unemployment and furlough figures caused by the pandemic. Source: NTB.

The need for these types of statistical summaries stemmed in part from a lack of clear data release strategy from government and health officials. According to a senior technologist at Bayerischer Rundfunk, health authorities were putting forward too many different indicators. He said that, for a time, they were communicating only about absolute increases and total numbers, which made it difficult to track the evolution of the virus, before switching focus to more detailed figures such as doubling times, reproductive numbers, and 7-day incidence rates. "There was a lot of confusion about which indicators are important: Should we look at deaths? Should we look at people in hospitals? Should we look at the number of tests conducted?" said the technologist. The decision to include some statistical indicators over others stirred up debate within the newsroom. For instance, in addition to important aspects that media organisations need to take into consideration when reporting on testing results (e.g., not being able to account for all the mild and asymptomatic cases that go unrecorded, comparing the number of positive tests against the total number of tests performed), he also raised a caveat that relates to the number of beds available in intensive care units:

> Sounds like a good indicator, but it's actually not because the number of beds doesn't tell you much because you need, like, staff, medical trained staff, to man those beds. And oftentimes, like, the hospitals were, like, totally overworked and they didn't have, like, enough personnel.

(Senior technologist, Bayerischer Rundfunk, Germany)

The team in charge of automated news at the Swiss media group Tamedia faced a similar set of challenges as they had to rely on local cantons instead of the federal government to get data. "The federal government wasn't really able to provide structured data in a machine-readable way and in a reliable way", said a senior computational journalist at Tamedia. The cantons collected and presented data in an accessible manner. That said, the downside of having to rely on such a wide range of local sources is that journalists need to keep a close watch at how they evolve. "All the journalists in Switzerland who provide automated data reporting for the COVID

crisis, they have to constantly think, rethink, and update their data sources", said the computational journalist.

These statistical roundups were also sometimes prevented from going into full-on automation mode because of a disconnect between generating text and integrating them into the news organisation's content management system (i.e., CMS), an interactive tool that is used for publishing online. At *The Times*, automated text (unlike visualisations) needed to be generated manually on a separate webpage, then pasted and copied into the newspaper's CMS after undergoing editorial check, a process described by the computational journalist at the newspaper as "a bit of a halfway house". This type of disconnect was seen as a hurdle, as the numbers might already be outdated by the time they were ready to be checked. The data and interactive team eventually got permission to republish the story just with new numbers, as long as they would still go through editorial check if the lead of the story changes, for instance when the epicentre of the pandemic moved from China to Europe.

Similarly, having to generate automated news separately on a webpage was also seen as a drawback at the Canadian Press. "It's annoying to do the copy and paste, and that's sort of the limit of where our technology is", said a senior computational journalist at the news agency. He stressed that this was be solved with the adoption of a new CMS: "We'll have API access to the content management system, so we won't even have to have these webpages", he said, indicating that automated news could even be generated using a simple command on Slack, a messaging program that is popular among journalists to coordinate teamwork.

5.1.2   Using automation to develop new workflows

Some newsrooms also used the pandemic as an opportunity to rearrange or come up with new types of workflows that involved either setting up automated news or working directly with it. At the British news agency RADAR, which provides automated coverage for regional titles across the United Kingdom, journalists contribute every step of the way to setting up automated news: this ranges from finding storylines in newly published datasets to authoring templates using a third-

party tool. An editor at RADAR described part of this process as looking out for datasets that possess a certain level of "granularity", meaning that they contain information at quite a local or hyperlocal level (e.g., local authority districts, police force areas), which journalists can use to find news angles to work with. Once they manage to get a sense of all the possible stories that can be told, they turn to the self-editing tool that RADAR subscribes to (i.e., Arria NLG Studio) so as to author templates for automated news. A sample of stories generated this way is then checked before all of them are released at once. In the case of Bloomberg News, automated news stories are connected to an machine learning system that extracts key information on companies' statements and analyses them through "knowledge graphs" (see Meij, 2019) that help prepare for various scenarios, using Bloomberg's internal data for this. These scenarios are then run live to produce automated news in multiple languages. Although this procedure involves machine learning elements at the event detection and analysis stages, writing scripts in advance to anticipate each of these scenarios remains largely a human effort.

In hindsight, these workflows proved to be robust enough to handle the data deluge that followed the spread of COVID-19. Usually, the team at RADAR would work on one or two projects a day, but the pandemic reshuffled the deck to four or five projects a day "just because there is so much data coming out", said the editor at the news agency. This briefly put the newsroom under pressure due to resourcing issues, but also gave the team an opportunity to showcase their expert knowledge of producing automated news—which resulted in more clients being interested in RADAR's other automated products. On its end, being part of a major company that specialises in technology and data gave Bloomberg News a head start and meant it had access to valuable resources: "We had a lot of datasets and a lot of procedures and a lot of technology available, so we were able to shift very, very quickly to the new environment, to the new topics", said an executive at Bloomberg. To report on the economic impacts of the pandemic, the news agency made notable use of alternate datasets to generate automated news:

> This is where we had all these very interesting datasets about subway or metro ridership, you can have booking for a restaurant reservation, you can have plane ticket reservation... So we have all these really interesting snippets of data where when you combine it for different countries, you can find a very

interesting sort of real time dataset that comes ahead of some of the country's economic data that really tells you the impact of the crisis locally.

(Executive, Bloomberg, United States)

The second type of newsroom workflow involved journalists directly working with automated news, using them as first drafts or to assist them in their own reporting. To provide journalists at the Finnish newspaper *Helsingin Sanomat* with fresh updates on the pandemic, an algorithm programmed by the newspaper's data team connected every morning to an API set up by the Finnish Institute for Health and Welfare (see also Piechota, 2020): "It goes to the newest COVID numbers and then creates short pieces of text and sends them to Slack to inform the reporters", said a computational journalist at *Helsingin Sanomat*, who was involved in developing the software (see Figure 13). She specified that journalists can publish these stories as-is or tweak them before publication. In the same way, journalists at the Swedish news service Omni turned to automated content provider United Robots to equip them with a breaking-news desk that delivered regular updates on COVID-19. United Robots' software connects to 24 curated sources via APIs and RSS feeds, or scrapes content associated with keywords on each of these websites (United Robots, no date, 2021). Every time new information is out, an alert with a link is sent on a Slack channel used by the newsroom (see Figure 13) so that journalists can include these details in their own reporting. United Robots regularly updates its sources based on Omni's suggestions.
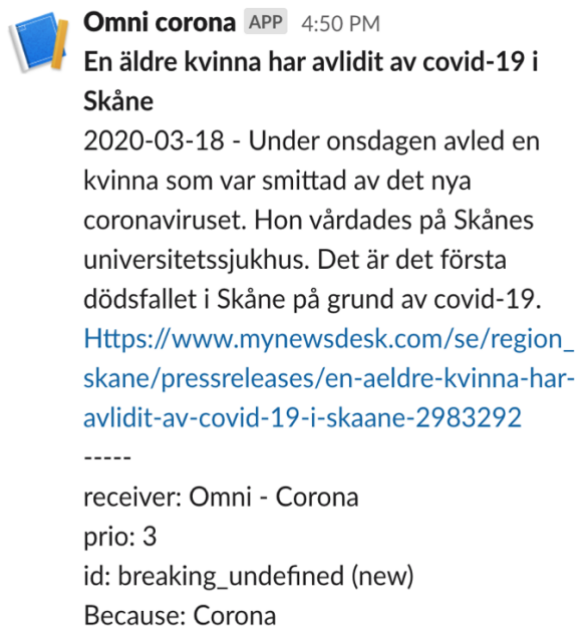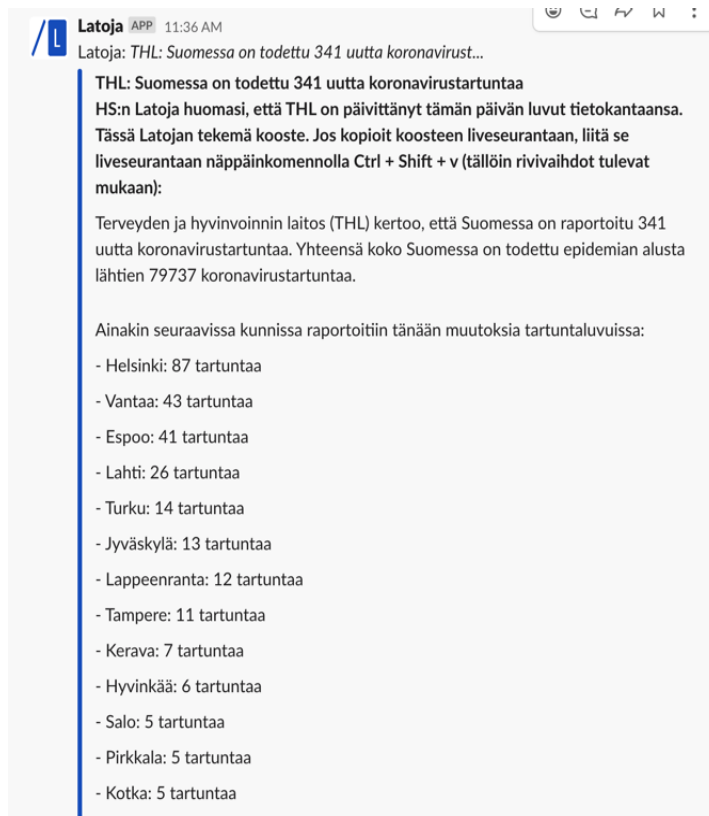
Figure 13: *Helsingin Sanomat*'s and Omni's use of automated news. The *Helsingin Sanomat* (top) and Omni (bottom) use an alert system through Slack that provides journalists with automated text to include wholly or partially in their copy, or to assist them in their own reporting. Sources: *Helsingin Sanomat* and Omni.

Both *Helsingin Sanomat* and Omni experienced difficulties in accessing local data initially, which led them to seek out algorithmic solutions to remedy that. "We didn't really have those kinds of extremely local news sources or health agency sources", shared a manager at Omni. Her only regret is that they adopted the dashboard a bit too late, once every local new case of COVID-19 was no longer considered breaking news. "For a long time, one single death somewhere was something that we wanted to send a push notification on", she said. Access to local sources was also considered to be an issue in Finland where, prior to the launch of the governmental API, reporters had to manually retrieve information from the 21 websites that represented each of the country's health areas. "Basically, somebody had to sit in front of their computer and refresh some pages to see if there are new numbers or not", said the computational journalist at *Helsingin Sanomat*.

In this section, I highlighted how twelve news organisations used automated news on COVID-19 to provide a statistical overview of the spread of the virus *via* user-facing interfaces and new media products, or turned to rearranged or new workflows structured around the use of the technology. Next, I will detail how automated news was used, this time, as part of a series of experiments at the BBC, which aimed at testing out its potential for news production.

5.2 BBC'S EXPERIMENTS WITH AUTOMATED NEWS

Unlike other media organisations that made the choice to outsource the production of automated news to external content providers or to build their own in-house automated news systems, the BBC made the decision to subscribe to an online platform, Arria NLG Studio, which lets journalists design their own templates for automated journalism: this is similar to RADAR's strategy (see section 5.1.2) or to other media organisations' use of self-editing tools like Automated Insights' Wordsmith in the United States or AX Semantics in Germany (Mullin, 2015; AX Semantics, no date 1). While basic data manipulation can be done within Arria, the team in charge of developing automated news at the BBC found that more complex programming needed to be done outside of it. Under the lead of BBC News Labs—the broadcaster's own incubator whose role is to test out new technologies for media

production—five experiments were conducted in just one year, each time resulting in further iterations that gradually became more elaborate: these were split between minor projects (i.e., accident and emergency waiting times, local elections, tree planting and high street shopping) and a more important one that was done as part of the 2019 United Kingdom general election. All of these were published online for BBC audiences, except for the high street project that was more of prototyping experience. These projects were all experimental by nature, hence the choice of the word "experiments".

5.2.1 From A&E waiting times to shopping on British high streets

The first automated news project started off as a feasibility check at the beginning of 2019, to see if the National Health Service data feed that was used to run BBC's health performance tracker could be turned into over a hundred automated stories each month, so as to inform readers on Accident and Emergency (A&E) waiting times in East Anglia trusts (Hutton, 2019). The News Labs team set up a system that fetched information directly from data already garnered by the Visual and Data Journalism team, which included the percentage of patients being taken care of within four hours of arrival and the last time the trust's target was met, or not, in a five-year period. Through the use of Arria Studio, the News Labs team prepared some narratives that could then be retrieved using a front-end interface (see Figure 14) and pushed to the BBC's content management system. They also arranged for automated visualisations to be created and displayed along each of these automated stories (see Figure 15).
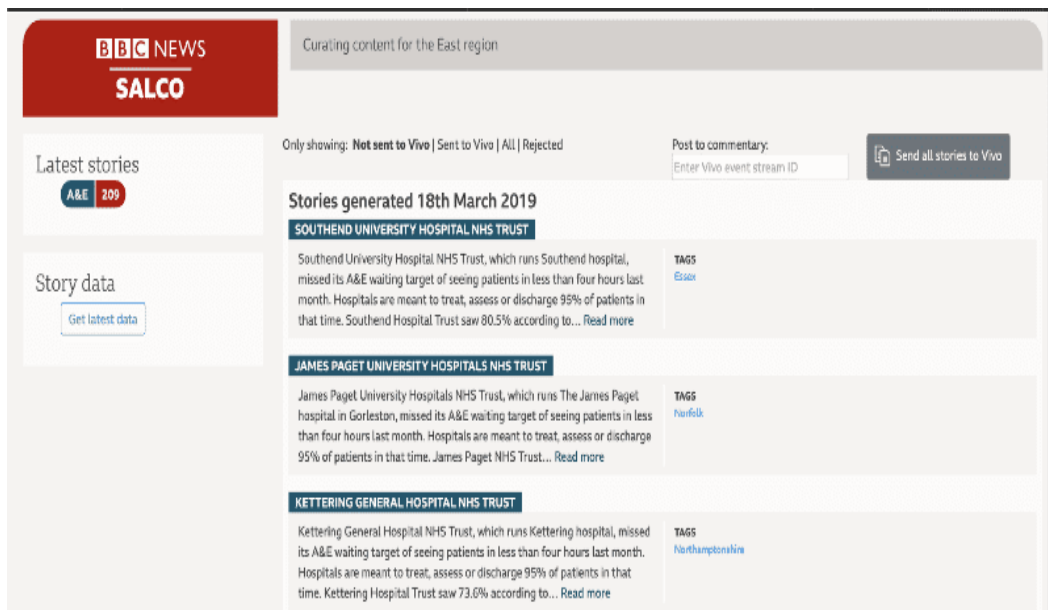
Figure 14: Front-end interface for BBC's A&E stories. The front-end interface that journalists used to look up stories generated on A&E waiting times in East Anglia trusts, which could then be pushed to the BBC's content management system. Source: BBC News Labs.
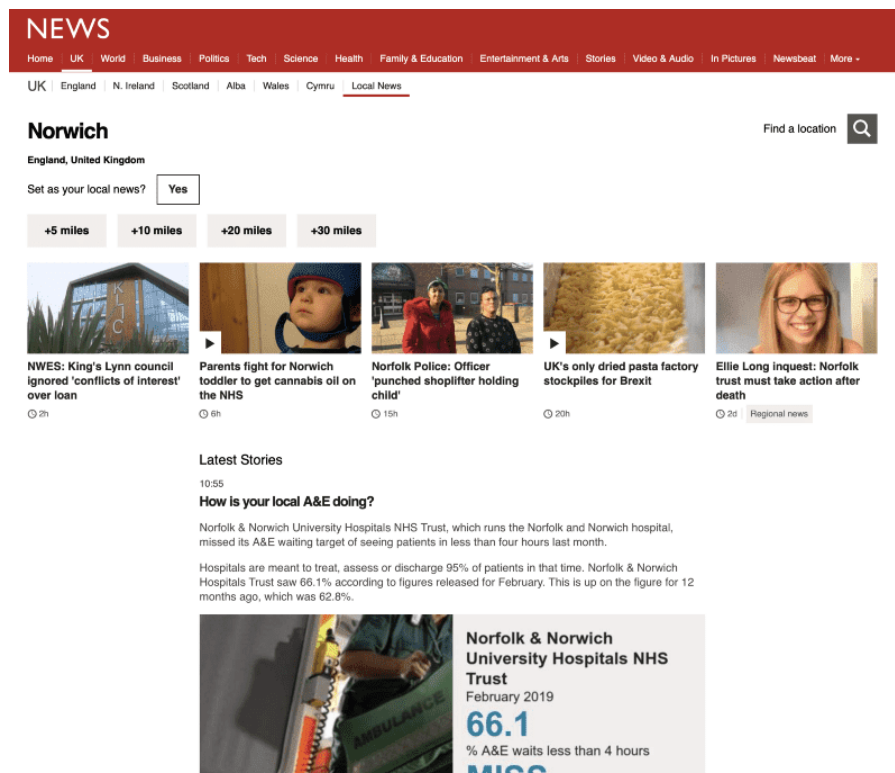
Figure 15: A&E story on a BBC local page. An automated story on A&E waiting times for Norfolk & Norwich University Hospital Trust, along with an accompanying automated visualisation, which was published on Norwich local news feed. Source: BBC News Labs.

The second project, an iteration of this initial version, took place shortly after, during the United Kingdom's local elections of May 2019 (Green, 2019). During this phase, the team tested out the potential to use the BBC's election results feed to generate automated news, focusing for that on a sample of 16 stories (see Figure 16): they wanted to see if the infrastructure they built as part of the A&E project could be linked to a live stream that takes results from a variety of sources, which include journalists at counting centres, news agencies and communications from councils administering elections (which are then manually inputted and/or verified within BBC's system). The team ran out of time to automate visualisations—like they did with A&E waiting times stories—but managed to show that their automated news system could be successfully employed to report on election results.

Figure 16: BBC local elections story. A story generated for Calderdale during the United Kingdom's local elections of May 2019. Source: BBC News Labs.

The third experiment conducted by the team at BBC News Labs looked at the use of automated news to produce over 300 localised stories on the rate of tree planting in England, which were generated as a complement to a larger data journalism piece published on 30 July 2019, showing that twice as many trees were needed to be planted in the United Kingdom so that it could meet its carbon emission targets (Molumby, 2019). This initiative was made possible through a collaboration between News Labs and BBC England Data Unit, drawing for that on data from the Forestry Commission as well as on estimates from the United Kingdom's advisory committee on climate change. All of these automated stories, which covered each local authority district in England and the city of London, would then link back to the main data journalism piece, which provided a nationwide overview (see Figure 17). The team also created additional templates using language suited to reading aloud, which covered different groupings of geographic areas. They distributed these to appropriate outlets within BBC Local Radio, so that they could use it as they saw fit. "That was interesting because it let the stations use the script, kind of put it in their own voices (...), which was kind of interesting because we want to make local and national news more joined up", said a senior BBC technologist. "If we're doing a story on a national level, being able to do a corresponding local piece that is entirely automated is kind of part of that strategy", he added.

How much more woodland should be planted?

Tree planting rates in the UK
Experts want tree planting back at levels last seen in 1989

30,000 hectares — The Committee on Climate Change says the UK should be planting 30,000 hectares of trees a year

About 13,400 hectares were planted in the year to March 2019

20,000 hectares

10,000 hectares

0

1979   1989   1999   2009   2019
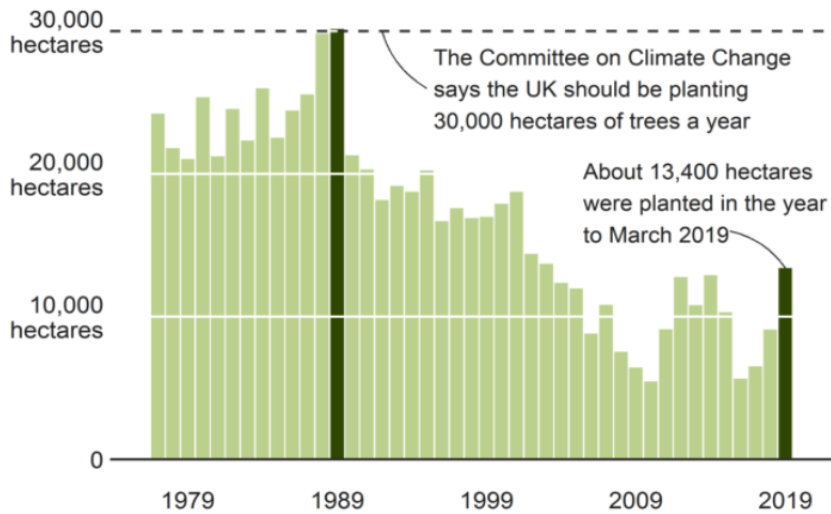
Note: A standard football pitch is about two thirds of a hectare - 1 pitch = 0.64 hectares

Source: Forest Research   BBC

The CCC said 30,000 hectares (116 sq miles) of new trees are needed per year until 2050.

This is equivalent to filling more than 46,000 standard football pitches or a space about three-quarters the size of the Isle of Wight every year.

- **Tree planting: Your questions answered**

- **Can I plant trees anywhere?**

Ewa Kmietowicz, the CCC's transport and agriculture team leader, said: "The government needs to develop a strategy to meet the 30,000-hectare target and it needs to happen quickly."

18:29 2 Aug

91,100 government-funded trees planted in Aylesbury Vale in eight years

There have been 91,100 government-funded trees planted in Aylesbury Vale between 2010 and 2018, Forestry Commission data shows.

This works out at 465 trees per 1,000 people.

Between 2010 and 2018 the government funded about 15 million trees in England. Figures for trees planted with private funds or by local councils are not included and the data only covers new trees, not replacements for any that were cut down.

Figure 17: BBC's use of automated stories with data journalism. The main data journalism piece on the rate of tree planting in England and an accompanying automated story generated for the district of Aylesbury Vale. Source: BBC News & BBC News Labs.

Soon after that, automated news was used for the fourth time in a lookup tool launched in October 2019, which gave access to nearly 7,000 hyperlocal stories trying to capture the extent to which people were still shopping on British high streets, given the popularity of online shopping (see Figure 18). To do this, a computational journalist managed to combine a geo-localised dataset from the Ordnance Survey, which listed up retail information on every high street in Great Britain, with employment figures from the Office for National Statistics. The journalist wrote computer scripts to compare the streets' numbers with economic data at the regional and national levels. He also created a separate set of about 350 stories that focused on high street retail at a local authority level, which included quotes that were gathered in advance from a business representative, who would provide two different types of answers depending on whether retail activity was higher or lower than average. However, these quotes did not figure in the final prototype, as the emphasis was rather put on the 7,000 hyperlocal high street stories.

> Based on which scenario matched the street, it would then use the correct quote in the template. (…) We believe it's the first time anyone's actually done that in a news story. (…) The key to that editorially was just being very transparent with the person that we were quoting in advance about how this was going to be used and why, and that it was still editorially valid, and also just, you know, assuring them that it is still experimental.
>
> (Computational journalist, BBC, United Kingdom)
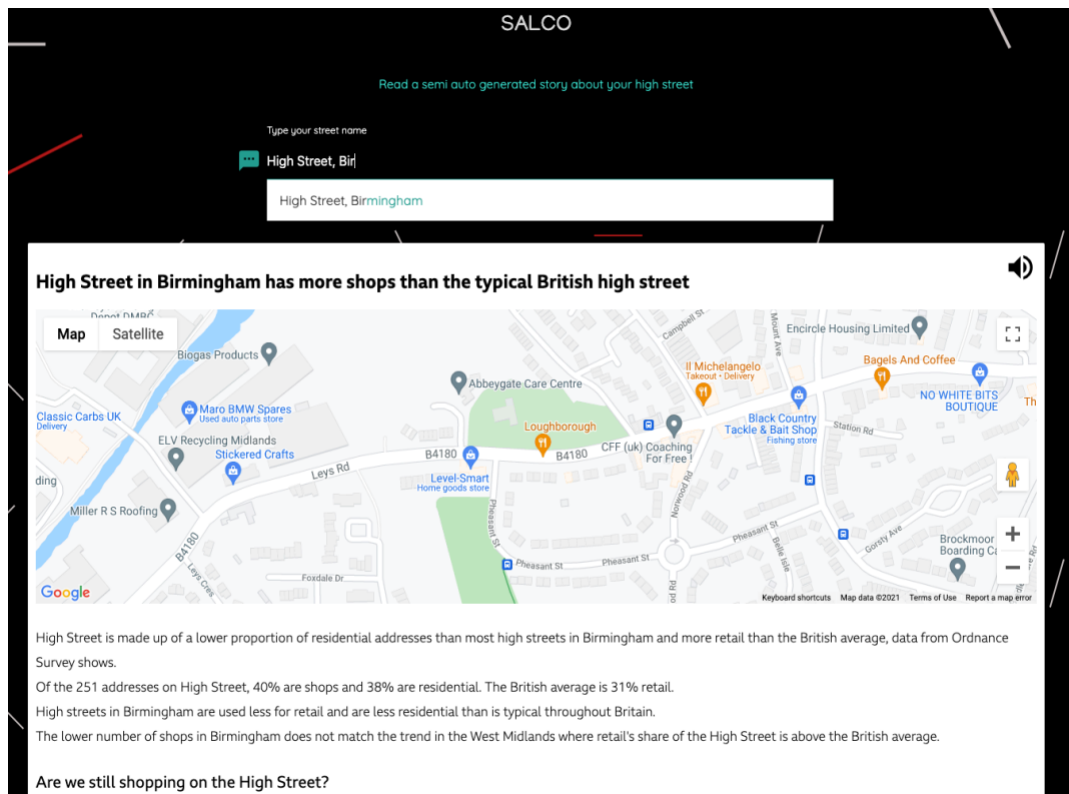
Figure 18: BBC's hyperlocal hight street project. The lookup tool developed by BBC News Labs, which gave access to close 7,000 hyperlocal stories trying to capture the extent to which people were still shopping on British high streets, for instance here in Birmingham. Source: BBC News Labs.

5.2.2 Covering the 2019 general election in the United Kingdom

The last BBC's experiment with automated news—by far the most ambitious—was the generation of close to 690 stories to cover the December 2019 general election in the United Kingdom, which included 40 stories written in Welsh (Molumby & Whitwell, 2019). For this, the News Labs team built on the infrastructure they already developed as part of their local elections coverage but, this time, reporters were also directly involved in the process as they were assigned with checking stories and adding missing details before publication. This team of a dozen journalists that spanned across the London, Cardiff, Glasgow and Belfast newsrooms was trained on NLG concepts in the weeks leading up to the election. A manager observed that human intervention was indeed necessary when dealing with edge cases in this election, as only journalists could "recognise the weirdness of the situation" and step

in. According to one journalist on his team, "there were moments of such complexity that basically we just gave that information to the journalist on the night and said, 'Just be aware of this, you might have to rewrite some of this because it's, like, weirdly complicated'".

To prepare for flagging these edge cases to the editorial team, information on seats that were considered to be "highly volatile" (i.e., likely to change side after belonging to the same party for decades) or "high profile" (e.g., the prime minister's or the opposition leader's seat) was gathered ahead of time with the help of the BBC's political research unit, which produces briefs and analyses for the news service. On the night of the election, these extra bits of information were provided to journalists through the Slack notifications they received to inform them when new stories to verify were ready (see Figure 19 and Figure 20). "We knew that we were dealing with journalists at 3:00 AM, 4:00 AM, 5:00 AM in the morning, and therefore you want to be as explicit as possible", said the BBC journalist. The original intent behind this was to help journalists move quickly through these edge case stories. In addition to this, they were also handed a checklist that helped them look up for potential mistakes, such as typos and phrasing that NLG could twist in an odd way.
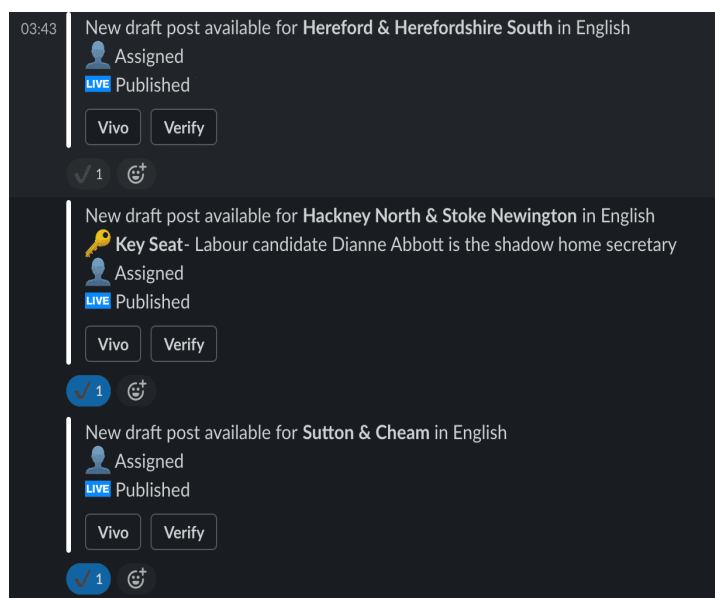
Figure 19: Slack notification used in BBC's general election coverage. An example of automated Slack notifications informing journalists that new results were in, which also provided them with extra information on constituencies that required further attention, for instance here the shadow home secretary's. Source: BBC News Labs.
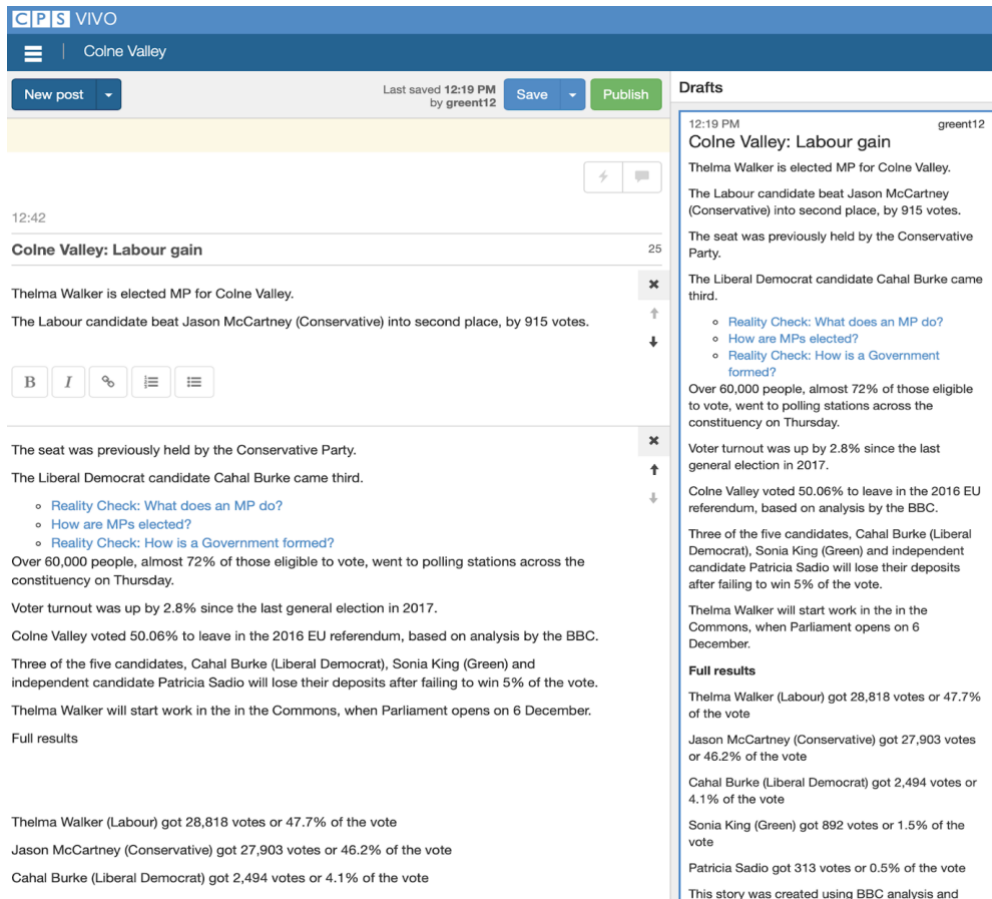
Figure 20: Automated story in BBC's CMS. The Slack notifications included a link to find the story in the BBC's content management system. From there they could make edits or publish as-is, as shown here in this test story for Colne Valley. Source: BBC News Labs.

This way of combining automated news with human input, also known as "combined journalism" (Wölker & Powell, 2021), proved to be successful with 690 stories being checked and published in about 10 hours. According to a BBC editor, ensuring participation from all of the BBC's national and regional newsrooms participated was vital to making sure that no important local context was missing. In Scotland, Wales and Northern Ireland, journalists used their expert knowledge of the political situation there to build on those automated pieces quite considerably before publication whereas, in England, the core London team was supposed to be prioritising speed of publication, leaving regional newsrooms to add details and local knowledge later.

The only mistake that was spotted on the night was introduced by a journalist who was not part of this project, who replaced an MP's first name by his last name following the "Sir" prefix, whereas automated news generated the right way of addressing a Sir by his first name. That being said, some of the journalists recognised that they lingered on some of the details more than what was necessary, as they started polishing up these pieces. This resulted in some automated stories having more details than planned, but also ended up delaying the overall publication schedule. For instance, a technologist (2) with a journalism background who contributed to editing these stories recognised that, in the early morning, she started being slower in processing the results because, as a journalist, she wanted to have "somewhat of an understanding of what this means for this constituency". Similarly, the BBC editor mentioned that he spent about 10 minutes adding in extra details on his own constituency—which he was knowledgeable about—instead of verifying and publishing additional pieces:

> I probably spent a good 10 minutes, um, adding in some local knowledge to it, which I shouldn't have done really, I could have checked off 10 other constituencies in that time. Um, so yeah, so I think that's the danger (...): where you put a human being into the loop, the temptation is for the human being to go, "Oh, I could make this even better".
>
> (Editor, BBC, United Kingdom)

I have described here five experiments that were run at the BBC in 2019, to test out the potential of using automated news: these included small experiments as in the A&E waiting times, local elections, tree planting and high street projects, but also a more ambitious attempt at covering the 2019 United Kingdom general election with automated news. In the next section, I will provide a more general overview of how automated news was implemented in different countries and media types (i.e., news agencies, newspapers, public service broadcasters).

5.3 AUTOMATED NEWS' IMPLEMENTATION ACROSS MEDIA ORGANISATIONS

As described in my methodology, I will conduct here a cross-national multiple case study using Hallin and Mancini's media system typology (2004) to strategically select news organisations—limiting myself to news agencies, newspapers and public service broadcasters—in order to reflect on their comprehension of *differentiation* and *de-differention* within the media industry, which mirrors Bourdieu's views on commercial homogeneity (see section 3.2.2). Additionally, when analysing how automated news is implemented within these organisations, I will make use of ANT to distinguish two types of strategies: first, using automated news as *intermediaries* where initial intent is kept and where it does what it is supposed to do; second, using automated news as *mediators* when something *more* is added to existing practices, in this case additional meaningful *translations* where new human and non-human *actants* get involved, which shows the overall direction that it is taking.

5.3.1   Predictable uses as *intermediaries*

First, based on some of the most prominent examples that are developed in section 1.2.2, it can be said that automated news is used as *intermediaries* when private or public service datasets are being used as sources, when there is no journalistic involvement other than through the affordances already provided for by third-party tools and—for now—when text only is generated, sometimes with visualisations. Such an assemblage can be observed at news organisations outsourcing automated news to external content providers, like at the Associated Press, where teams collaborate with firms like Automated Insights and Data Skrive to come up with templates so that these companies can automate corporate earnings stories and sports recaps[31], based on private data (see Colford, 2014). Likewise, Italy's news agency ANSA publishes weather forecasts that are sometimes generated using automation and data provided by a weather forecast company, but also national and regional accounts of the spread of COVID-19, using public data collected through Narrativa's

---

[31] The data team at the Associated Press is also involved in setting up automated news for smaller scale projects, like polling results for each of the 50 American States or when transforming an exclusive dataset into local stories.

COVID-19 tracker initiative[32] and put together by the firm Applied XL (Redazione ANSA, 2020; Narrativa, no date 2). As for Spain's national public service broadcaster RTVE, it collaborated with Narrativa to run trials on less watched football competitions in Spain using private data and also prepared for generating stories on election results in small municipalities based on government data (Corral, 2021). This is similar to what the French public radio broadcaster France Bleu and French newspapers belonging to the Belgian media group Rossel (e.g., *La Voix du Nord, L'Union*) have been doing during recent elections in France with automated news generated by the firms Syllabs (France Bleu) and LabSense (Rossel), based on governmental data. Besides, the Belgian newspapers group Sudpresse (owned by Rossel) and LabSense also collaborated on automating amateur football games in Belgium, sourcing data from a sports association.

Automated news used as *intermediaries* is also visible when it is designed internally. As such, Reuters' data team has been developing automated news the usual way while setting up stories on sports, financial news and COVID-19, relying both on private and public data. This was also true of *The Times*' automated journalism project on COVID-19, which was based on public data and programmed in-house. As for the Norwegian news agency NTB, it relied on a select few editorial developers with both a journalistic and technical background so as to be able to automate the same type of pandemic-related content as well as sports, election and financial news (see Figure 21), using private and public data for this.

---

[32] The Spanish public service broadcaster RTVE was also part of this initiative, mostly as an information provider, but is not mentioned here as it did not necessarily used automated news produced this way in a systematic manner.

Figure 21: NTB's automated football story. An example of an automated football game recap that was generated at NTB. Source: NTB.

On their end, the data team at the newspaper *Stuttgarter Zeitung* programmed automated news to cover the 2021 German election at a municipal level with local governmental data, while a team of technologists at the Finnish public service broadcaster YLE developed automated summaries on sports and election results, using both private and public service data and helped by a journalist who can understand code. Moreover, YLE made its code for generating ice hockey recaps open source, following a Parliament's request to limit unfair competition in the Finnish media market: as a result, other organisations like Finland's news agency STT used this code for their own ice hockey stories. Sometimes, an academic partner was also involved in the development of automated news, as in the Bavarian broadcaster Bayerischer Rundfunk's collaboration with the Technical University of Munich to automate match reports for a basketball league in Germany (Sebis Research News,

2021; Schneider & Köppen, 2021), which came in parallel with the broadcaster's other project on COVID-19 and led to automating financial results as well (Schneider, 2022). To do this, the team relied on public health sources for the COVID-19 newsletter and on private data for sports and financial news. Lastly, after experimenting with their own solution to automate the Rio Summer Olympics, the 2016 presidential election in the United States and high school American football coverage (WashPost PR Blog, 2016a, 2016b, 2017), the *Washington Post*'s engineering team joined forces with Northwestern University to develop a "computational political journalism R&D lab" ahead of the 2020 presidential elections (Schmidt, 2019), this improving existing automated news models that draw on data collected by private brokers during election time.

Using automated news as *intermediaries* can also be found in the use of third-party self-editing tools that feature a form of No-code language, which allows editorial staff with little programming experience to design automated news on their own. This could be observed at the BBC, where the News Labs team used Arria NLG Studio to template out articles on A&E waiting times, tree planting and high street shopping, using public service datasets. The Swiss newspaper group Tamedia used Wordsmith—Automated Insights' own NLG technology that was made directly accessible to clients through a self-editing interface (Mullin, 2015)—to draft out automated stories on referendum and election results in Switzerland (Plattner & Orel, 2019; Marchand, 2019) and to provide a statistical roundup of the spread of COVID-19, using public service datasets. As for the Australian public service broadcaster ABC, it subscribed to a bot-building application, Chatfuel, to create a messenger bot (see Figure 22) that uses public service data to inform users on electoral results (Elvery, 2016; Archer, 2016), but also to provide them with daily news summaries, weather forecasts and emergency alerts (see Ford & Hutchinson, 2019).
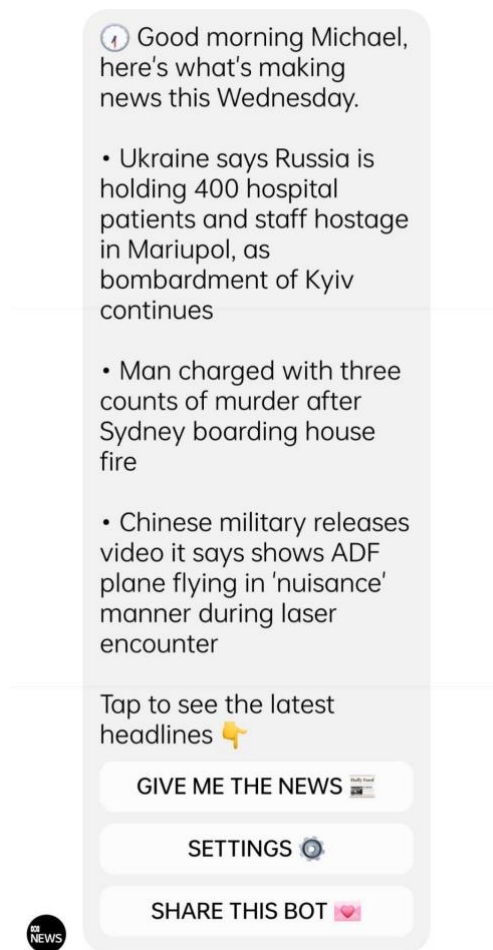
Figure 22: ABC's chatbot. A daily news brief delivered by the Australian Broadcasting Corporation (ABC) through its conversational chatbot platform. Source: ABC.

5.3.2 Transformations as *mediators*

In contrast to automated stories being used as *intermediaries*, mapping their roles as *mediators* requires carefully thinking about additional meaningful *translations* that new human and non-human *actants* could bring: this could be whenever these changes concern using sources other than private and public service data, deploying systems that are specifically built for journalists—other than through the affordances already provided for by third-party self-editing tools—and, lastly, generating outputs other than text. First, with regard to additional sources, a noticeable *translation* occurs as news organisations turn to their own internal feed, proceed to their own

data collection or use archival material, thus avoiding the need to rely on third-party private or public service datasets. An example of this is the BBC's and ABC's efforts to connect their automated news system to an internal election results feed (see section 5.2 for BBC), which in the case of ABC is linked to the corporation's own psephologist:

> We're mostly looking at the data sources we use for broadcast to start with, or that are at that level. (...) The election one is coming from the Australian electoral commission or the State electoral commissions, but then it's going through our election expert's system, Antony Green. So it's being processed by his system and he's taking those raw figures and putting his knowledge of electoral systems over them to come up with predictions and things like that.
>
> (Manager, ABC, Australia)

In a few instances, news organisations collected data on their own in order to automate news text, as shown in AFP's and Reuters' statistical roundups on the spread of COVID-19, which were both automated using shared spreadsheets that were manually filled by journalists on the ground, even if at Reuters this system was also connected to open data sources. As for tapping into archival material, the Finnish news agency STT collaborated for a time with the University of Turku, in Southern Finland, to automate ice hockey recaps using machine learning models (Kanerva *et al.*, 2019) that were trained on STT's own archives that dated back to the 1990s (see Figure 23). That being said, an executive at STT indicated that content generated this way did not meet the agency's standards to be delivered to clients, but was accessible to them should they be interested in it:

> It [STT's archives] goes way back, but there wasn't enough reports for the AI because (...) it just wants all the data and more and more and more... And it wasn't enough for the AI to learn enough. And the second thing was that there was too much human... well, too much human in them. So there was, like, adjectives and things that weren't in the data that the machine was fed. (...) For example, in ice hockey there was, like, standings that weren't in the data that the machine was given. So we ended up using some manual work to go through, not all, but a lot of the stories.
>
> (Executive, STT, Finland)
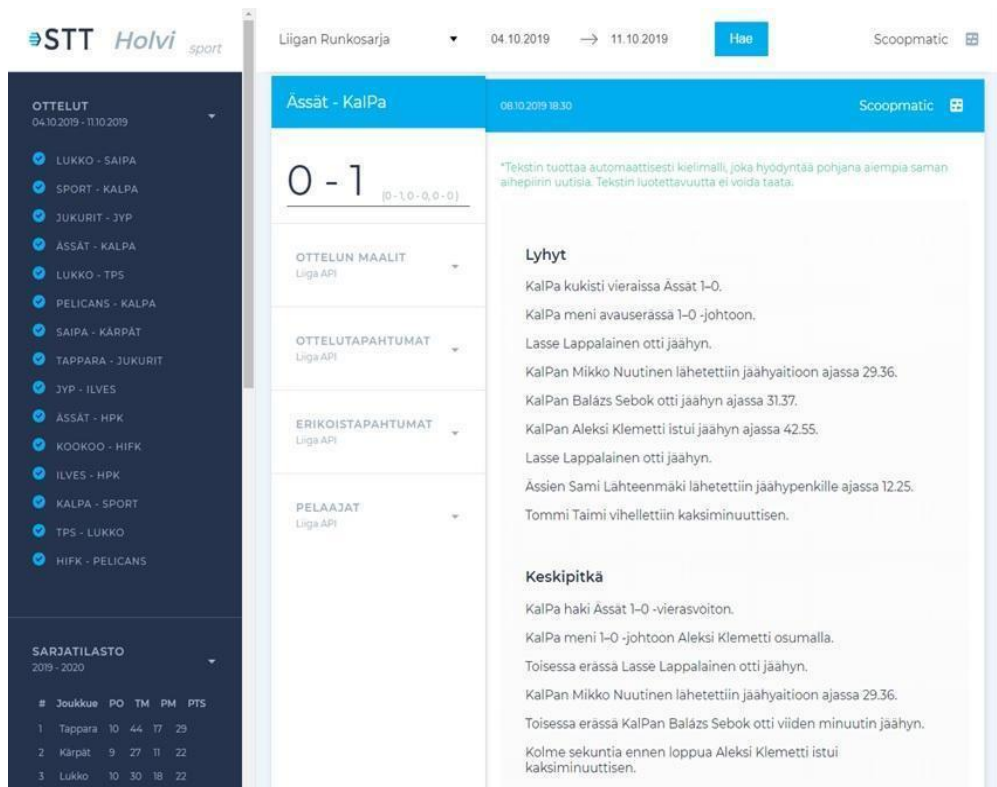
Figure 23: STT's machine learning stories. The interface on which ice hockey stories were generated using machine learning models that were trained on STT's internal archives, which covered games that were held since the 1990s. Source: STT.

Additional *translations* that related to source type were also visible in situations where crowdsourced material and social media feeds were used. The German newspaper *Stuttgarter Zeitung* relied on crowdsourced material to automate its air quality reports in the Stuttgart area, which were generated using AX Semantics' self-editing tool and connected to open data sourced from a network of community sensors (Plavec, 2017; Toporoff, 2017). In Australia, the ABC used opinion data collected through a polling exercise that is habitually done during election time, so as to come with answers for its messenger bot (Gee & Prior, 2018), an approach that was further extended to probe the public's concerns on emergency preparedness. Social media feeds, on their end, were put to contribution using web scraping techniques, so as to be able to collect user-generated content and to conduct computational analysis on it. This was done, for instance, at the Spanish public service broadcaster RTVE, which partnered with the University Carlos III of Madrid to

generate automated football stories that adopt a tone and voice that reflect users' opinions (del Rey García, 2020). "You can have the version for one team, for example: 'It was a great success,' the balanced news, and, on the other hand, (...) 'they stole us the football match'", said an executive at RTVE. Likewise, Reuters' News Tracer acts as a "breaking news radar" while roaming on Twitter feeds to find relevant information, using advanced detection, classification and evaluation techniques for this; it then goes on to generate short automated text that is passed on to journalists for verification (Liu *et al.*, 2017; Emerging Technology from the arXiv, 2017).

Another area where additional *translations* are brought into force relates to automated news systems that are specifically built for journalists, other than through the affordances already provided for by third-party self-editing tools. These can be, first, internal software that comes with its own self-editing tool, features notification streams and provides access to auto-generated background information, as in Reuters' Lynx Insight system where journalists can template out their own stories using a form of No-code language that resembles those of third-party tools (see Figure 24), receive Microsoft Teams notifications once stories generated this way or that the data team set up are ready (similarly in that sense to the BBC's Slack notifications in section 5.2.2) and query this system as they look for automated news with background information on the subject that they are covering.
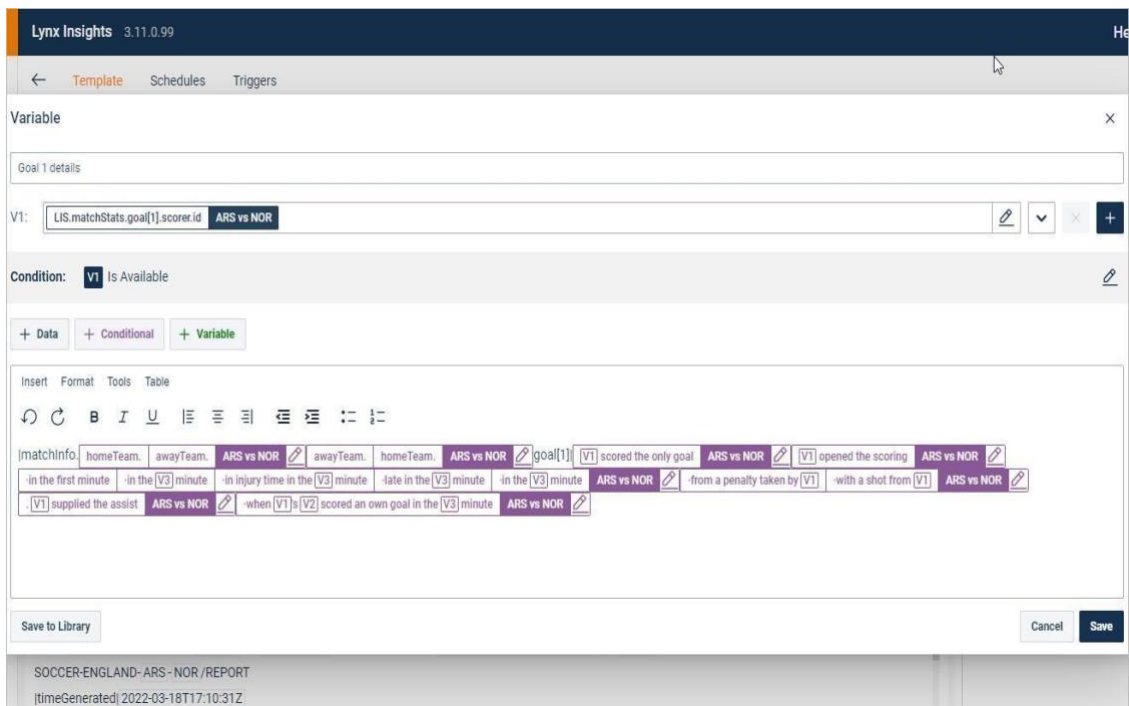
Figure 24: Reuters' Lynx Insight platform. On Reuters' Lynx Insight platform, journalists can template out their own automated stories using a form of No-code language that makes it accessible to news staff with little programming experience. Source: Reuters.

The online newspaper *El Confidencial* constitutes another example of the use of an internal self-editing tool, as the data team set up previews that help journalists visualise the automated story they are about to generate, instead of having to work right off computer scripts. "We prepared a web tool that they could use [that includes] the new text [created] with a different condition, and in real time they can see (...) how the final article will look like", said a technologist at the online newspaper. In another example of the use of automated backgrounders, the engineering team at the *Washington Post* and Northwestern University teamed up to create a query system that lets journalists access automated background information on the 2020 presidential election in the United States (see Figure 25): this contained for instance indications on the number and ethnic distribution of new registered voters in a given county (Diakopoulos *et al.*, 2020; WashPost PR Blog, 2020a).

Figure 25: The *Washington Post*'s automated backgrounders. At the *Washington Post*, a query system was set up in collaboration with Northwestern University so that reporters could access automated backgrounders that would help them cover the 2020 presidential election in the United States. Source: Diakopoulos *et al.*, 2020.

Finally, one last *translation* that could be observed has to do with generating output other than text, in this case NLG-to-audio content (see section 1.2.2). This was visible in the *Washington Post*'s and ABC's efforts to create stories of their own, and not just audio content suited to help with vision impairment. Automated audio stories generated this way could then be tailored to a listener's location as in the *Washington Post*'s election updates that were inserted in the newspaper's political podcasts (WashPost PR Blog, 2020b) or be accessed *via* virtual assistants (e.g., Amazon's Alexa) as in the ABC's diffusion of emergency alert summaries that were created using its own NLG tool (Collett, 2021, 2022).

In this section, I highlighted how automated news is being used at 18 news organisations that were selected based on Hallin and Mancini's (2004) media system typology (i.e., North Atlantic, North/Central and Mediterranean models) and that featured different media types (i.e., news agencies, newspapers, public service

broadcasters). Using ANT, I demonstrated that some of them used automated news as *intermediaries*, where initial intent is kept and where it does what it is supposed to do, while others considered it more as *mediators* while bringing in additional meaningful *translations*, which go as follows: first, enrolling alternate data sources while relying on a media organisation's own internal feed, data collection or archives as a source or, else, on crowdsourced material or social media feeds (see also Bloomberg's use of alternate datasets in section 5.1.2); second, enticing journalists while putting them at the centre of interfaces that are designed internally like in-house self-editing tools, notification streams and/or search features to access automated backgrounders; and, third, enlisting vocal elements through generating NLG-to-audio content as an output. These *translations*, which are summarised in Table 6, will be further examined in chapter 7 to see how they relate to *differentiation* and *de-differentiation* theory, but, before doing so, I will move on to the more interpretative part of my research where I address media practitioners' reactions to automated news.

Table 6. Automated news as *mediators*: areas with additional *translations,* per media system.

NORTH ATLANTIC

		News agencies	Newspapers	Public service broadcasters
Sources	Internal feed	–	–	ABC/BBC
	Own data collection	Reuters	–	–
	Archives	–	–	–
	Crowdsourced material	–	–	ABC
	Social media feeds	Reuters	–	–
Interfaces	In-house self-editing tool	Reuters	–	–
	Notification stream	Reuters	–	BBC
	Automated backgrounders	Reuters	Washington Post	–
Outputs	NLG-to-audio	–	Washington Post	ABC

NORTH/CENTRAL

		News agencies	Newspapers	Public service broadcasters
Sources	Internal feed	–	–	–
	Own data collection	–	–	–
	Archives	STT	–	–
	Crowdsourced material	–	Stuttgarter Zeitung	–
	Social media feeds	–	–	–
Interfaces	In-house self-editing tool	–	–	–
	Notification stream	–	–	–
	Automated backgrounders	–	–	–
Outputs	NLG-to-audio	–	–	–

MEDITERRANEAN

		News agencies	Newspapers	Public service broadcasters
Sources	Internal feed	–	–	–
	Own data collection	AFP	–	–
	Archives	–	–	–
	Crowdsourced material	–	–	–
	Social media feeds	–	–	RTVE
Interfaces	In-house self-editing tool	–	El Confidencial	–
	Notification stream	–	–	–
	Automated backgrounders	–	–	–
Outputs	NLG-to-audio	–	–	–

6   Media practitioners' own interpretations

After my more descriptive account of how automated news has been employed across a wide range of organisations and in different media contexts, I will now provide a more interpretative and critical examination of the way media practitioners reacted to it. This will help me determine the perceived impacts automated news has on the work of media practitioners, which will be summarised and discussed in chapter 7 so as to answer the research questions that I have set out. As a guiding thread here, I will use the three themes that were identified at the end of my theory chapter (i.e., over-reliance on external datasets; media practitioners acquiring a computational thinking mindset; conflicts within and outside the field of journalism). This second empirical part will come as a cross-case analysis of the three cases that were previously discussed (i.e., COVID-19, BBC and cross-national case study) as I draw on evidence that was collected in those different set of circumstances. Unlike chapter 5—which is more descriptive and so features a large array of netnographical elements (i.e., screenshots and material published online)—this chapter heavily relies on interviews as it is more interpretative and critical.

6.1   Challenges of having to rely on external datasets

The first of these themes relates to Bourdieu's concept of structures external to the field of journalism (see Wu, Tandoc & Salmon, 2019b) as it deals with having to rely on external datasets to automate news text. As shown in chapter 5, this is mostly due to the over-reliance on public service datasets (e.g., governments, health authorities, specialised agencies) or on private ones (e.g., data brokers, sports leagues), unless media organisations proceed to their own data collection, make use of their own internal feed and archives or, else, draw on crowdsourced material and social media sources. In this section, I will delve into the use of public service and private data to generate automated news, and unveil challenges that have to do with releasing strategies and data quality issues. I will then focus on some of the solutions implemented to address those.

6.1.1 Experiencing data releasing issues

One of the main issues that media practitioners had to face when dealing with external datasets to automated news relates to releasing strategies. This became evident during the COVID-19 pandemic, as they had to rely on multiple levels of government to get data, sometimes even coping with uneven times of release. The news agency Canadian Press experienced this with regard to Canada's regional governments. "It's hard to get this data from multiple governments: they all put it up in a different fashion, they put it up at different times", said the senior computational journalist. He stressed that governmental data published in Canada did not necessarily come as a machine-readable format, and that important information was sometimes shared at press conferences and in news releases, which made it harder to use for automation:

> It's not like: "here's the data, here's the daily data, here's a CSV or an API endpoint" where you can, you know, you can just take it. (…) You know, we could write scrapers for that, but the other thing is (…) they will do a press conference and say: "okay, there's 200 new cases". It'll be six or seven hours later before that appears on a government website somewhere and, you know, we just can't, we can't wait seven hours to publish these stories.
>
> (Senior computational journalist, Canadian Press, Canada)

In the United Kingdom, *The Times* grappled with the same lack of cohesive efforts in releasing data: "England, Scotland, Wales, and Northern Ireland all publish it in different ways", said the computational journalist at the newspaper, adding that this made comparisons between regions even more difficult. His team also had to come to grips with Public Health England releasing COVID-19 data quite unevenly at the end of the day, generally between 4 p.m. and 7 p.m., but sometimes as late as 10 p.m. (Watts & Joiner, 2020). For its part, the news agency NTB faced an even bigger challenge: "Every Norwegian municipality has a different way of providing the data on their own pages. So it's like 356 municipalities, and they have 356 different ways of sharing the data", said an executive at NTB. Outside of COVID-19, the absence of

data normalisation between public bodies was also seen as a challenge at the Australian broadcaster ABC, especially when it came to automating news on natural disasters (e.g., fires, floods), which are reported at an emergency management level in the country: "All that information splits into different States, but the emergencies don't recognize our state boundaries", said a manager at the broadcaster, who added that this pushed them to aggregate and normalise this data into one place (see Fell, 2021a). Reflections around data source selection also came into play when working with multiple levels of government. The senior computational journalist at the Swiss media group Tamedia pointed out that, at the beginning of the outbreak, they needed to choose between data released by some of the Swiss cantons, which, in some cases, accounted for deaths in elderly care that were likely to be caused by COVID-19, and data from the Federal Office of Public Health, which focused only on laboratory-confirmed deaths:

> Some of the cantons (...) they count people who die in elders' homes and who don't have a COVID test. They still count them (...) as [coronavirus] deaths, for some reason, whereas the federal government doesn't do that. So the number by the federal government is always a bit lower. So (...) that's a decision we need to take, (...) which figure to use, right? And here I think, you know, there's no right or wrong... but what you need to do is you need to be very transparent about (...) which dataset you use, right?
>
> (Senior computational journalist, Tamedia, Switzerland)

Sudden changes of format were also seen as a hurdle to efficient data releasing strategies during the pandemic. According to the computational journalist at *The Times*, the data structure used by Public Health England appeared to change very often. "It was really difficult to rely on them as a source", he said. The news agency NTB faced similar problems with data being released by municipal governments in Norway. "The sources might start reporting it differently or make adjustments. That's a challenge", said an editor at NTB. The data team at the Finnish newspaper *Helsingin Sanomat* experienced similar difficulties when working with data from the Finnish Institute for Health and Welfare: unlike morning data, which is dispatched *via* a central and structured API, afternoon data is released in an HTML table format (see Piechota, 2020): "It's terrible the website where this info is published, it's not any API

or anything. It's just a... It's just an HTML table there, (...) they keep changing the location of it. So they [the data team] cannot just take it automatically, but somebody needs to check that the numbers are really correct and then copy-paste it ", said the computational journalist at the newspaper.

Another challenge of having to rely on external datasets relates to the data itself—in other words data quality—starting with missing context that does not necessarily figure in the data source. This lack of contextual elements is one of the main drawbacks that explains AFP's decision not to pursue a pilot project it had with the French NLG firm Syllabs, in order to automate football news stories. As a senior journalist specifies, these stories based on a data stream were missing out on key details that make for a good depiction of the game:

> We tried to automate France's and then European football championship results in several languages. We thought "That's good, we can produce content at scale". What we realised after a time is that delivering results is fine, but it's not what a reader really wants to see. (...) What they want to see, it's: "Why did the referee whistle at that moment?", "Why did this player get tackled?", "Why was there no whistle on that?" It's all these emotional aspects that we're interested in.
>
> (Senior journalist, AFP, France, translation)

The same concerns about missing context also drove the Norwegian news agency NTB not to include weather information in its automated football coverage: "If you, say, use the forecast and then saying that, you know, it was a beautiful Sunday afternoon, (...) but in reality there was a hailstorm coming in that wasn't (...) forecasted and the match was postponed for 15 minutes, you know, it should have been mentioned, but it wouldn't be available in the data", said the editor at NTB.

In addition to missing context, null or missing data could also be seen as an issue when automating news text. The executive at the Spanish broadcaster RTVE mentioned the difficulties they faced when trying to automate football games with a score of 0–0: "So it's very complicated to create a news from zero. So even in [these] cases, they [Narrativa] are doing a piece of news just based (...) on the data that they have. I think that (...) the next step is to add more information, to do it more complete

in the sense that you have different opinions, different (….) points of view", he said, referring to RTVE's partnership with University Carlos III of Madrid to automate football stories using user generated-content (see section 5.3.2). The senior technologist at the Bavarian broadcaster Bayerischer Rundfunk said he saw problems like this happening at other organisations during COVID-19: "In the area of soccer where, because of the (…) COVID-19 crisis, there were, like, some games that were cancelled, but, like, the bots still did report on them in a very weird way or… yeah, there were, like, very weird scenarios", he raised. As a matter of fact, he mentioned the only time Bayerischer Rundfunk's COVID-19 newsletter would fail is when the feed provided by the Robert Koch Institute would stop working:

> The only scenario which would (…) break it, it's (…) when we are not getting data from the federal agency. And that has happened before, but usually we… Well, people will tell us [laughs] because they like, uh, our product and somehow they depend on it and, well, if it doesn't work I get a call. (…) No matter what time of the day.
>
> (Senior technologist, Bayerischer Rundfunk, Germany)

6.1.2   Working around having to rely on external datasets

To tackle some of these challenges of having to rely on external datasets, more or less developed computational solutions and strategies were designed internally within newsrooms. First of all, some of the newsrooms under study reported having "computational safeguards" in place, this so as to make sure that erroneous content does not creep into the final version of automated news products. At the *Washington Post*, such safeguards were visible in scripts generating an error message whenever a situation was deemed to be too unrealistic: this could be for instance if the total number of votes would go above 100% or formatting issues such as when something other than a number would show up where only a number should be, according to the newspaper executive. At Reuters, an editor mentioned having computational safeguards in place to avoid pushing automated news with errors like missing data or illogical mathematical operations (e.g., divisions by zero).

In addition to these computational safeguards—designed to deal with automated news published at scale—other computational solutions were found to address the specific challenges of having to rely on external datasets. Similarly to Reuters' and AFP's spreadsheet system where journalists manually input new COVID-19 numbers for the location they are based in (see section 5.3.2), the Canadian Press resorted to the same kind of technique to get data from every regional government in Canada, and across different time zones:

> So I built a simple Google sheet, that's got the different provinces, the number of cases, which confirmed cases, and then also presumptive results, yes. (...) And it does all the, you know, all the maths to add them together (...) All that the people have to do, like, all the reporters have to do is go in and type in the new numbers in the page (...) You put your information in, you go back to the page, you click on generate story and it automatically generates a story.
>
> (Senior computational journalist, Canadian Press, Canada)

At the *Times*, a dedicated computational solution was used to deal with uneven times of release: to avoid journalists having to frequently refresh Public Health England's webpage, the team set up an alert algorithm that checks in the database available online if the date corresponds to that of today's (Watts & Joiner, 2020); if so, a Slack notification is sent to inform journalists that new data is out, so that they can generate their own automated news and visualisations using the webpage that was built for this purpose. A similar solution was found at the *Helsingin Sanomat* to address sudden changes of format: the data team set up an alert algorithm that sends a Slack notification every time numbers on the Finnish Institute for Health and Welfare's HTML table are updated, which informs journalists they need to retrieve the numbers manually and enter them into a shared spreadsheet that is used to generate automated stories. In spite of this, the computational journalist still observed that—as the Institute's database changes tended to be made at the end of the working week—this would sometimes result in the data team working overtime on Friday evenings, so as to make sure that there would be no technical glitches over the weekend:

We are already leaving and then we notice that "Oh my God, they have changed their systems" and then we need, like, [to] do it in the Friday evening, which is [sigh] frustrating. But I try to believe that they are not doing it [on] purpose: they are just humans and they want to have things done before the weekend comes and then they leave it at [that last minute].

(Computational journalist, *Helsingin Sanomat*, Finland)

Yet, the most comprehensive framework to address both releasing strategies and data quality to date rests with the BBC's "building the story model" approach, where journalists are being asked to critically investigate any data source they use. "A naive approach to natural language generation is to assume the data is somehow, you know, useful or accurate. The data might not be accurate, and it might not be useful and it might be biased", said the BBC manager. He stressed the importance of having a data "quality control" step in the automated news design phase, in order to make sure that data meets the organisation's standards, and also to act as a safeguard prior to writing templates: "It's not data, template, story", he observed. "It's data, *story model*, template, story."

Essentially, the BBC's "story model" step consists in running analysis on one or several datasets to determine which elements are newsworthy enough to include in automated news templates, and then gathering them into a table (see Figure 26). A story model step has for instance been at play in the high street project, where the computational journalist analysed a database that compiled information on every high street in Great Britain, so he could find the most interesting scenarios to work with (see section 5.2.1). He shared that getting familiar with the data this way helped him have a clearer picture of what was at stake:

So say I got a trove of data from... uh... some power supplier and I wanted to see... um... where is, like, the least efficiently consuming power or where is the most, and so on. I would just query that dataset to try and build what I would think will be a new story out of it. You basically just have to [detach] yourself further from that dataset and think "What are the possibilities within this dataset?" and write the logic accordingly.

(Computational journalist, BBC, United Kingdom)

That being said, there are also a few caveats associated with building the story model, one of them being doing wrong associations. As such, a BBC journalist warned against the risk for readers to infer meaning just because that two data pieces are put next to each other:

> Ultimately (...) you are asking the reader to infer meaning from the fact that those two pieces of information are next to each other and to a degree you're then telling a story. And whilst that's absolutely fine in journalism—that's the point—(...) if you were to do that 600 times, sometimes it would just be not appropriate to put those two pieces of information next to each other. And (...) the story that you're kind of suggesting is probably not valid or not useful or not accurate.
>
> (Journalist, BBC, United Kingdom)

health_trust_code	name	region	nation	date	ae_figure	ae_over_4_hours	ae_12_months_ago_figure	ae_target_achieved	ae_rank	idt_uuid
RJ1	Guy's & St Thomas' NHS Trust	London	england	01/01/2019	87.5	2255	92	Jul-15	32 of 131 trusts	548c0638-36a1-11e9-bea6-06a3ab30b7c4
RQX	Homerton University Hospital NHS Trust	London	england	01/01/2019	92.5	829	94.2	Nov-18	11 of 131 trusts	548c11be-36a1-11e9-bea6-06a3ab30b7c4
RYJ	Imperial College Healthcare NHS Trust	London	england	01/01/2019	86.7	3382	85.1	Jun-15	38 of 131 trusts	548c1b96-36a1-11e9-bea6-06a3ab30b7c4
RJZ	King's College Hospital NHS Trust	London	england	01/01/2019	71.9	6921	77.7	Jun-15	116 of 131 trusts	548c2546-36a1-11e9-bea6-06a3ab30b7c4
RAX	Kingston Hospital NHS Trust	London	england	01/01/2019	86.8	1475	87.3	Apr-16	35 of 131 trusts	548c3054-36a1-11e9-bea6-06a3ab30b7c4
RJ2	Lewisham & Greenwich NHS Trust	London	england	01/01/2019	83.8	4247	83.3	Jun-15	59 of 131 trusts	548c3a68-36a1-11e9-bea6-06a3ab30b7c4
R1K	London North West Healthcare NHS Trust	London	england	01/01/2019	85.6	4656	83.6	Not in the last five years	43 of 131 trusts	548c4404-36a1-11e9-bea6-06a3ab30b7c4
RAP	North Middlesex University Hospital NHS Trust	London	england	01/01/2019	80.9	3135	75.5	Jul-15	78 of 131 trusts	548c4da0-36a1-11e9-bea6-06a3ab30b7c4
RAL	Royal Free London NHS Trust	London	england	01/01/2019	85	3691	86.1	Oct-15	47 of 131 trusts	548c58cc-36a1-11e9-bea6-06a3ab30b7c4
RJ7	St George's University Hospitals NHS Trust	London	england	01/01/2019	84.2	2413	83	May-15	58 of 131 trusts	548c629a-36a1-11e9-bea6-06a3ab30b7c4
RAS	The Hillingdon Hospitals NHS Trust	London	england	01/01/2019	80.7	2789	80.8	Sep-15	79 of 131 trusts	548c6c36-36a1-11e9-bea6-06a3ab30b7c4
RRV	University College London Hospitals NHS Trust	London	england	01/01/2019	83.3	2026	86.1	Jul-15	62 of 131 trusts	548c75c8-36a1-11e9-bea6-06a3ab30b7c4
RKE	The Whittington Hospital NHS Trust	London	england	01/01/2019	86	1343	86.5	Sep-15	40 of 131 trusts	548c8112-36a1-11e9-bea6-06a3ab30b7c4

Figure 26: BBC's "story model". The "story model" used to create the A&E waiting times stories. In this table, information needed to calculate performance targets was selected. Source: BBC News Labs.

In addition to the "building the story model" step, other methods and techniques were used at the BBC to ensure data quality before automating content. The senior technologist recommended having an institutional knowledge of any data source that is used regularly. "It's not entirely unusual with some of these datasets that they are withdrawn or they are amended and republished", he raised. In the BBC's first automated news project on A&E waiting times, the team would also let newly published data "breathe" for a bit, before performing a series of extra checks on it. One of these checks involved for instance a deduplication procedure, whereby an algorithm was run against the most recent database to check whether new data is any

different from those already published, so as to avoid pushing the same automated story twice.

Another illustration of these checks involved troubleshooting procedures, such as when a technologist (technologist 1 in Appendix B) reported finding issues after merging two datasets that contained some electoral candidates' details: "There could have been some instances where a candidate was matched by the number and they weren't the same candidate, or there could have been situations where there's a candidate with the same name, but they weren't the same person", he said. Likewise, his computational journalist colleague experienced similar issues while running a test based on historical data:

> One of the things that we had managed to spot was that we had built in, into our logic, to pick up whether or not somebody had retained their seat, so if it was the same MP (...) our headline would be, you know, conservatives' hold or labour hold as in the party has held the seat because it's the same MP. But what we actually discovered was that a couple of MPs changed their name on the register between elections. And the best example I found, the reason I found it was the Brexit secretary, Steve Barclay, was down as Stephen Barclay in 2017, and Steve in 2019. So it said conservative gain in this test headline when actually it was a hold.
>
> (Computational journalist, BBC, United Kingdom)

In this section, I have shed light on some of the challenges that relate to having to rely on external datasets to generate automated news. I have demonstrated that, on the one hand, some of these issues concern data releasing strategies (i.e., having to deal with multiple levels of government, uneven times of release and sudden changes of format) and that, on the other hand, data quality considerations also come into play (i.e., missing context, null or missing data). To address these challenges, more or less developed computational solutions and strategies were implemented within newsrooms, such as computational safeguards to deal with automated stories being produced at scale, a shared spreadsheet system to get data from multiple levels of government as well as across times zones, and algorithmic alert systems that let journalists know when new data is out or that warn them of any sudden changes of format. Having said that, the most comprehensive framework to date lies with the

BBC's "building the story model" approach, where journalists are being asked to critically question any data piece they wish to include in automated news templates. In light of all these, it is also important to remember that having to rely on external datasets echoes other challenges found in mainstream journalism practice, which deal with having to work with a limited number of authoritative and official sources for routine news, thus ultimately hampering diversity in news coverage (Gans 1979; Molotch & Lester 1974). In the next section, I will focus on another core aspect of the use of automated news: the need for media practitioners to acquire a computational thinking mindset.

6.2   DEVELOPING A COMPUTATIONAL THINKING MINDSET

The second key area to look at relates to Bourdieu's concept of acquiring cultural capital (see Wu, Tandoc & Salmon, 2019b), in this case the need for media practitioners to develop a computational thinking mindset (Diakopoulos, 2011; Stavelin, 2013; Gynnild, 2014), understood here as a way to solve problems through applying a form of abstract reasoning close to computer programming (see Wing, 2008). In the section below, I will first detail how a "structured journalism" approach is employed to help media practitioners better handle abstraction concepts, then go into the specifics of embedding a media organisation's own polices into computer scripts.

6.2.1   Structured journalism to work with abstraction concepts

Setting up automated news very much requires a computational thinking mindset as media practitioners need to address editorial considerations in the code they are using (see Dierickx, 2021), relying for that on a type of abstract reasoning that is used in computer programming. In practice, this is made possible while taking a "structured journalism" approach, a process of "atomizing the news" (Jones & Jones, 2019) so that narratives can be turned into databases (Caswell & Dörr, 2018; Anderson, 2018). As raised by the BBC manager, this change in mindset implies thinking about stories no longer as individual pieces, but rather as "patterns that emerge":

The kind of the skill (...) that's kind of central to writing those templates is basically the ability to work with stories abstractly, instead of just in terms of the specific story, right? So, some journalists they just, you know, (...) they think in terms of the specifics, not in terms of the patterns that emerge, whereas when you're creating those templates, you know, it's still writing, (...) but you're doing it at the level of the pattern of stories, of all the possible stories, (...) not just at the level of the specific, that's the key, the key thing.

(Manager, BBC, United Kingdom)

The BBC editor suggested on his end that—rather than writing a story that is perceived as "your beautiful piece of work that's completely owned by you and [has] nothing to do with anybody else"—drafting out templates for automated news could be compared, in a way, to a storytelling technique known as "the story spine" (see Figure 27), which consists, first, in listing out a set of recurring elements such as "Once upon a time..." and "But, one day...", and then use them as prompts to come up with the story's specifics (Adams, 2013).

THE STORY SPINE

THE STORY SPINE	STRUCTURE	FUNCTION
Once upon a time...	Beginning	The world of the story is introduced and the main character's routine is established.
Every day...		
But, one day...	The Event	The main character breaks the routine .
Because of that...	Middle	There are dire consequences for having broken the routine. It is unclear if the main character will come out alright in the end.
Because of that...		
Because of that...		
Until finally...	The Climax	The main character embarks upon success or failure
And, ever since then...	End	The main character succeeds or fails, and a new routine is established.

Figure 27: Example of a "Story Spine". The "Story Spine" model created by playwright Kenn Adams, which lists out recurring storytelling elements that can then be used as prompts to come up with the story's specifics. Source: Adams, 2013.

In real-world applications, adopting a structured journalism approach to set up automated news works best for baseline scenarios, first by foreseeing what an ideal story would look like, and then breaking it down into smaller elements that can be reusable across many versions of that same story. "If you didn't have any automation involved, what would be the story you, as a human being, would want to write, or what would be the elements of the story that you would want to write?" raised the BBC editor. "Having established that, we then looked at what data we could get to fill that". As a demonstration of this process, the BBC computational journalist described searching for all the possible elements that could be thought of advance, in preparation of the 2019 general election in the United Kingdom: "You don't know in advance who is going to win the national general election, you don't know who is going to win each seat, but you know all the possibilities in advance", he said. These could include, for instance, who could win the race, how close the victory margin could be as compared to last election, the possibility of a dead heat, whether it could be a gain or a loss for each party, the other candidates' ranking as well as whether they may lose or not their deposit. Once all these possibilities have been set out, they constitute a starting point for scenario planning and contribute to building the logic that goes behind the construction of each template. The computational journalist explained that this process eventually leads up to elaborating six to seven fundamental sentences that are regarded as the "bare bones" of the automated story (see Figure 28), which can be reusable across many versions of it.

Florence Eshalomi has been elected MP for Vauxhall, meaning that the Labour Party holds the seat with a decreased majority.

The new MP beat Liberal Democrat Sarah Lewis by 19,612 votes. This was fewer than Kate Hoey's 20,250-vote majority in the 2017 general election.

Sarah Bool of the Conservative Party came third and the Green Party's Jacqueline Bond came fourth.

Voter turnout was down by 3.5 percentage points since the last general election.

More than 56,000 people, 63.5% of those eligible to vote, went to polling stations across the area on Thursday, in the first December general election since 1923.

Three of the six candidates, Jacqueline Bond (Green), Andrew McGuinness (The Brexit Party) and Salah Faissal (independent) lost their £500 deposits after failing to win 5% of the vote.

This story about Vauxhall was created using some automation.

Figure 28: BBC general election story. An automated story generated for Vauxhall constituency during the 2019 general elections in the United Kingdom. It includes details that could be thought of in advance, such as by how many votes the leading candidate has won and how it compares to last election, voter turnout and which candidates lost their deposit. Source: BBC News.

The same process of envisioning what an ideal story would be like and then breaking it down into reusable elements could also be seen at other media organisations. Hence, a similar "working backwards" approach was mentioned by an executive at the Associated Press: "What does the story need to look like? And what are all of the possibilities? You know, earnings go up, earnings go down, earnings stay flat. You know, I mean there's all of the branches that follow depending on the data that you're using to produce the automated story", she said. This thought process was also at play at the Bavarian broadcaster Bayerischer Rundfunk when creating templates for automated basketball stories: "We usually start, like, with an ideal article for, like, one case, where we would for example write a perfect basketball match report and then we kind of try to *templatise it*, make it into a template, and find out what's possible and what's not possible", said the senior technologist at the broadcaster. This whole business of abstraction took an even bigger turn at the

145

Washington Post, where—during the 2020 presidential election—the engineering team partnered with Northwestern University to probe journalists as to what type of details they would like to see included in automated backgrounders. "We worked really closely with our newsroom and we asked them: what are the things that you're interested in? What are the heuristics that you would use if you were looking over this dataset, you know? And I think generally that was, like, a pretty good strategy too because (...) we probably would have vastly over-thought it", said the executive at the newspaper. "We would have been, like, trying to characterise every county by, you know, after its rural, urban, suburban divide or, like, try to get all this other stuff out of it." To carry out this task, the engineering team collaborated with a computational journalism scholar, who conducted interviews and made a few prototypes in order to figure out what media practitioners consider to be newsworthy:

> The work that was particularly hard that he needed to do (...) was essentially trying to figure out how a reporter or an editor arrives at a heuristic for what is interesting or what is newsworthy, right? And he did that through a series of interviews, through, like, um... multiple, like, prototypes and try rounds. You know, we finally came up with something that, like, captured essentially the process that our reporters and editors would go through.
>
> (Executive, *The Washington Post*, United States)

Having said that, some media practitioners seemed to have had a more difficult time than others to be able to come to grips with this process of abstraction. The BBC senior technologist noted that, in some of the trainings that were held to familiarise reporters with concepts of NLG production, some participants could easily engage with what he described as a "complex tree of a story", whereas others were reluctant to thinking in advance of all the possible permutations:

> To kind of composite blocks that may be combined in different ways, um, some people couldn't kind of reason what that story would be and that became more difficult (...) whereas other people (...) were instantly really engaged with the idea that "Oh, if this happens, it could have this whole kind of branch of the story that only exists under certain conditions". That seemed to be kind of the distinguishing thing: some people were kind of in tune with that, um... kind of complex tree of a story and other people were much more reluctant to do that.

(Senior technologist, BBC, United Kingdom)

In the same vein, an editor at the German newspaper *Stuttgarter Zeitung* remarked that, during a workshop held by AX Semantics to explain how its self-editing tool worked, some of his colleagues had a harder time comprehending the abstraction concepts at play:

> There were three, four colleagues, who were interested in (...) [and] also participated (...) in this initial workshop. They wanted to know "how does it work?", but you could... They talked themselves and you could see on their faces that they kind of dropped out after one, two hours, because it was too complicated for them and they, um... they had real problems to (...) um... to think like, like, um... like a computer would do, like a program would do.

(Editor, *Stuttgarter Zeitung*, Germany)

To palliate this, some newsrooms reported having changed or adapted their recruitment policies so as to make sure that news staff involved in automated news or computational journalism projects do possess these abstraction skills. As a matter fact, the BBC manager specified that journalists at News Labs are partly hired based on their comfort to work with abstraction, and have to sit a small test before joining the team. Having journalists who are comfortable working with numbers and abstraction onboard is especially seen as an asset at RADAR, as the news organisation's media clients do not necessarily have this type of expertise in-house, nor have the time to invest in it. "We have to recruit people who are very comfortable working with numbers, much more comfortable than the average journalist", said the editor at RADAR, who himself worked as a business analyst before going into journalism. "So people who... you know, we don't have sort of data scientists or statisticians as such, but we do need people who are very comfortable working with numbers [and] understand what's going to be happening with that data." The Norwegian news agency NTB pushed these prerequisites even further: part of its recruitment strategy is now focused on hiring journalists with a programming background. "So we tried also to teach other journalists, especially the template

coding, but it's... it's easy to get a developer to understand journalism than the other way around", said the editor at NTB. His executive colleague stressed for his part how important it was to have the "right people" on the team:

> It's quite easy if you have the right people with the right, um... with the right *heads* [laughs]. And of course it's... I mean, it might be like 80% of the programmers would have never been able to understand journalism and those who will, um... it's really, it's much, much better to work with them than to try to teach a journalist coding.
>
> (Executive, NTB, Norway)

6.2.2   Embedding journalistic knowledge into code

Another aspect that is linked to the acquisition of a computational journalism mindset has to do with embedding a media organisation's own standards and practices into code for automated news. According to the BBC manager, doing so required being very specific about these rules:

> If it's an editorial requirement that you can deal with in writing then by definition, because of the way these tools are structured, (...) you can deal with it in the template. The challenge is—and we came across this very much in the (...) sort of the lead up as we were preparing for the election—the challenges is in articulating very specifically (...) what those editorial rules are (...) and a lot of them are written down, like in (...) the policy guides and the style guides and all the rest of it. But some of them are not, you know.
>
> (Manager, BBC, United Kingdom)

This could involve, for instance, reflecting on the right choice of words to qualify an electoral win: "If it's by, you know, 50 votes, you know, then you might call that, you know, very narrow win or whatever; if it's by, you know, 50% of the votes you might call that an enormous win", said the manager. "You've got to figure out where the boundaries are for the words that you use." On the night of the general election in the United Kingdom, the implementation of a "combined journalism" or "human-in-the-loop" form of workflow showed that, on this aspect, there was still room for improvement. The technologist with a journalism background (technologist 2 in

Appendix B) who contributed to vetting these stories on the night said that, sometimes, she felt the need to correct some of the headlines, as those were only mentioning a win with over 50% of the votes when, in fact, the winning party secured over 75% of votes:

> It wasn't, like, a major concern, I just kind of was, like, "this isn't as accurate as it could be" and I think it's important to reflect, like, that. (…) I just was aware of, like, how complicated election coverage is and how much we're… how frequently the BBC is accused of bias? So I just didn't want, like, an under-reporting of the margin of victory to be, like, taken as bias.
>
> (Technologist [2], BBC, United Kingdom)

Editorial issues like these could be found across media organisations using automated news to cover election or referendum results. The editor at the Norwegian news agency NTB remembered having to decide on which small parties to report on individually—and not for example under the label "others"—which implies making a call as to which threshold to use:

> In Norway we have a lot of parties in the elections, especially local elections: we have small parties into every city and, you know, (…) single cause parties, etc. (…) We had one on toll taxes on the roads, with a party against that which started as a local party and then started in (…) more cities. They were still quite small in many places, but when should you decide that they should be reported separately and one shouldn't. So all those kinds of judgements you have to make based on the data, which is quite hard sometimes.
>
> (Editor, NTB, Norway)

Similarly to BBC News Labs' collaboration with political experts to delineate edge cases ahead of the 2019 general election in the United Kingdom (see section 5.2.2), editorial staff at the *Washington Post* were asked to contribute their political expertise so as to come up with potential "edge cases" within the United States' electoral system:

> For those really odd outcomes where things go to a runoff, for example, or where there is no winner declared on election night or a variety of edge cases like this, we needed a ton of extra help from reporters and editors to essentially figure out what those edge cases were and then how we would like to handle them using *Post*'s style. And in some cases we actually dug in and found, you know, like, really bizarre situations that (...) we didn't ever end up having to prepare for but we were ready for it just in case. Like, for example, if the number of electors was 269 on each side and there was a tie in the electoral college, we had a brief ready for that.

> (Executive, *The Washington Post*, United States)

Likewise—in the case of a double majority referendum that, in Switzerland, requires the support of most citizens and also at least half of the cantons—the senior computational journalist at Tamedia mentioned having to watch out for unlikely outcomes, like when a proposal is backed by popular vote, but not by a majority of cantons:

> That's an edge case because that happens very rarely, right? Usually if the popular vote is above 50%, so if the majority accepts it, usually the majority of cantons will also accept it. And so when developing these templates, I could have forgotten about this edge case, right? And then the text would have said "well, the vote was accepted because (...) 52-53% (...) of the people accepted it", but that text would have been plain wrong because it would still have been refused.

> (Senior computational journalist, Tamedia, Switzerland)

Aside from articulating those very specifics in editorial decision-making, designing automated news also required having to rethink journalistic codes of conduct. For instance, in the BBC's high street shopping project (see section 5.2.1), the computational journalist indicated that he managed to get quotes in advance from a business representative, who provided him with two types of answers based on whether retail activity was higher or lower than average. "So basically they were able to give me a quote for both scenarios. So places where the high street had more kind of activity or retail than before (...) or where it was declining", he said. "And based on which scenario matched the streets, it would then use the correct quote in the

template." According to him, this approach is best suited to generating performance-oriented stories, like those that aim at determining whether hospitals, police or waste collection services meet their targets. "You have a certain number where, you know, they have met the target or they haven't met the target. So you say, 'If they haven't met the target, what are the typical reasons for that?' And they would give you their analysis", said the computational journalist.

However, getting quotes in advance could also be frown upon, as this would raise editorial questions. The senior BBC technologist alluded to the team's first experiment with automated stories, where a professional association—like a doctors' union—could have been asked to comment on A&E waiting times:

> So you may go to a union representing doctors and say "what would you say if this target wasn't met?" and "what would you say if the target was met?" And then we can include those quotations, but that did raise some... an editorial... difficulties about how do you include a quotation that's attributed to a person if they haven't actually responded to the thing that happened? They basically hypothetically responded.
>
> (Senior technologist, BBC, United Kingdom)

That being said, the computational journalist specified that, according to him, this is editorially valid as long as journalists are being transparent and explain to their interviewees the logic behind gathering quotes in advance for automated news:

> It's just making sure that you're not going to misquote the person by putting it in in the wrong scenario or context. So basically that you are going to do what you've told them you're going to do with that quote, or why that quote is relevant, you know. So that you don't throw it in somewhere where it's irrelevant and it basically looks like they don't know what they're talking about.
>
> (Computational journalist, BBC, United Kingdom)

At last, another aspect of encoding journalistic standards and practices into code has to do with to a news organisation's own style and tone. The executive at the Associated Press stressed that, as such, there can be "no wiggle room" when it comes

to embedding the specifics of the news agency's stylebook into computer scripts for automated news. This is all the more important as the Associated Press' guidelines act as a reference not only for the news agency, but also for the rest of the American news industry:

> We need to be able to have a template that conforms to how we need that story to read: I mean, everything from the dateline, you know, the location at the front of the story, you know—a little "AP" and parentheses, you know—like, I mean, all of that stuff, you, um... I mean, we have master lists of what we call every company on first reference, on second reference, you know, do you shorten the name of the company on second reference? Like, all of that is in columns of data that tells the template what language to use.
>
> (Executive, Associated Press, United States)

Similar concerns were raised at *The Times* and France Bleu regarding rounding up numbers rather than giving exact figures (*The Times*), and avoiding using the same phrasing twice in the title and lead (France Bleu). At Sudpresse, debating on stylistic issues brought a healthy discussion with the firm automating content for them: an executive at Sudpresse remembers opposing LabSense's suggestion that—when a football team would lose zero to five—the corresponding story would read as if the team has been "crushed": this, he said, could be interpreted as a "pseudo-editorial" decision. Moreover, according to Belgium's football rules, a score of zero to five can also be attributed in situations where one of the teams forfeits the game:

> At some point, LabSense came to us with suggestions and we declined them as they were pseudo-editorial. I'm explaining myself: for instance, LabSense was saying that when a team loses 5-0 then it's been "crushed" by the other team. But we didn't really want to go that way because it's just data and we don't know what happened on the playing field. So we didn't want this to backfire. (...) So we decided to remain quite "cold" and neutral.
>
> (Executive, Sudpresse, Belgium, translation)

Encoding a media organisation's own style and tone into a programmatic form touches on language issues as well. As an organisation that operates in multiple

languages, the BBC also produced Welsh-language automated news during the 2019 general election. In preparation for this, an assistant editor at BBC Wales sat with two members of the News Labs team to adapt English-written templates into Welsh. Often, the trio came to realise that an extra programming command was needed for Welsh: this could involve, for instance, having to change the order of elements in a sentence or modifying the beginning of a word depending on whether the noun is feminine or masculine, similarly in that sense to French or Spanish. This last point turned out to be troublesome for the team, as the data source they were using did not include the candidates' gender. While templates could be written gender-free in English, it created an extra layer of complexity in Welsh. "Whether the candidates, the winning candidate, was male or female, we would have to construct the sentence differently", specified the assistant editor.

To the same extent, the executive at Sudpresse observed that, sometimes, additional work was necessary to adapt some of the phrasings suggested by LabSense—which used French from France to generate automated news—into a language that would better suit the francophone community in Belgium. As for the automated football news project that RTVE and Narrativa were working on, the executive at the Spanish broadcaster said he asked the start-up not to use too vernacular terms:

> For example, how to say "referee" (...), just use the specific word because we are not a sports media, so we don't use the kind of style or words that uses a specific sports media. So we are more, um... wide in the sense of news: we don't give you opinions, we don't give you strange words or names: it's just a match and this is how it happens.
>
> (Executive, RTVE, Spain)

I have illustrated here how media practitioners and newsrooms alike used a structured journalism approach to implement the type of computational thinking mindset that is needed in automated news projects. It involves, first, considering what an ideal news story would look like, and then breaking it down into smaller pieces that can be reusable across stories. If this prompts newsrooms to be on the lookout for news staff that are well-versed in concepts of abstraction and in dealing with

numbers, it also turns out to a problem for those who see a story only in terms of its specifics and not as "patterns that emerge". Having said that, the type of skills required to be able to think of a story in an abstract way resembles, in a sense, those that are at play when writing "prep copies" ahead of time, like in the case of "advance obits" (see Adams, 2020). Acquiring a computational thinking mindset also implies having to rethink journalistic standards and practices as those are translated into code for automated news. In the last section of this more critical and interpretative chapter, I will focus on the types of tensions that stem from the introduction of automated news within newsrooms.

6.3   CONFLICTS WITHIN AND OUTSIDE THE JOURNALISTIC FIELD

To conclude this chapter, I will now focus on the types of tensions that take root in conflicting views or agendas that relate to the implementation of automated news. These can between media practitioners and, on the one hand, businesspeople and technologists working in news (see Lewis & Westlund, 2015b) and, on the other end, players external to the journalistic field like Big Tech companies. This corresponds to Bourdieu's conception of the field as a "site of struggle", as media practitioners may adversely react to increased algorithmic automation within newsrooms (see Wu, Tandoc & Salmon, 2019b). In this section, I will first examine the types of tensions that occur between editorial staff and technologists, then explore those at play between newsrooms and platform companies. At first glance, it may seem like tensions that concern the business side are missing: this is because it is harder to document critical views towards people that sit in a position of power (here executives, who are representative of the business side at a managerial level) and also because I was unable to interview people from the marketing and advertising teams: this is due to news organisations rather dispatching editorial staff, executives or technologists to talk about their automated news projects, instead of members from these business branches (whom I did not have access to because I was unable to conduct newsroom ethnographies as a result of COVID-19). That said, tensions that concern the business side could be perceptible in criticisms of managerial ways of organising journalistic work.

6.3.1 Adversarial logics within the newsroom

Even though no outright tensions between editorial staff and technologists could be observed in plain sight—mostly, again, because I was unable to conduct newsroom ethnographies—these were nonetheless visible in a more covert way. A few statements made by some of the technologists seemed, indeed, to testify of their frustration with regard to editorial staff's handling of abstraction and numbers. This was perceptible in the way some Canadian Press journalists entered new COVID-19 numbers in the shared spreadsheet that was set up to automate the coverage of the disease:

> Basically people won't read instructions (…), like, this really hammered it. This is very simple: you go into a spreadsheet, you change the number; that's all you've got to do. I'm not asking you to, you know, to do any calculation, to do any math: click on a cell, type in the new number and it's good. (…) Don't copy and paste, don't drag and drop, (…) write in the new number, that's all you have to do. And I show people how to do it and still multiple times we've got, like, "Emergency! Emergency!" The spreadsheet is all screwed up, somebody copied and pasted in something that wasn't a number and then all the calculations stopped working or, you know, people decide they're going to put commas or decimals and things and it's like… you know, why would you need a decimal? It's an integer. (…) Half a person doesn't get sick or whatever.
>
> (Senior computational journalist, Canadian Press, Canada)

To a similar extent, the senior technologist at the Bavarian broadcaster Bayerischer Rundfunk indicated that he did not necessarily approve of editorial managers' own preferences towards COVID-19 indicators, especially when it came down to important aspects of reporting on testing results (e.g., comparing the number of positive tests against the total number of tests performed) and to the number of hospital beds available in intensive care units, which were not necessarily manned by enough medical staff (see section 5.1.1):

> I'm now dealing with this crisis for… seven months! (…) Sometimes, like, being strong and, like, kind of doing your own thing was, like, really important and really helped the quality of the product. (…) When we are talking about, like,

different indicators and what indicators are important and... that's something we, as data journalists, could, like, much better evaluate.

(Senior technologist, Bayerischer Rundfunk, Germany)

In contrast, editorial staff rather positioned themselves as expert sources on journalistic matters and routines. This was palpable in the way editorial staff at *The Times* acted, in a sense, as gatekeepers while not letting their data and interactive colleagues update an automated copy with new numbers:

> The copy is not really changing other than the numbers, but it would still need to be manually entered into the system. So to make a change, it was kind of... justify why we were making a change. We kind of have to say (...) we're not actually changing the story, the copy isn't going to read any differently, it's just the numbers that's going to change (...) So we kind of had to convince them that... Well, we had to basically adapt to them, not being able to make the changes as frequently as we wanted (...), but we eventually kind of set the team up so that we can make those changes ourselves.
>
> (Computational journalist, *The Times*, United Kingdom)

In addition, the computational journalist at the Finnish newspaper *Helsingin Sanomat* remarked that editorial staff and technologists had a different understanding of transparency when it came to correcting content, as journalists were keener on having a disclaimer featured:

> That's maybe one mental difference I have noticed between journalists and coders: that coders just think that this is okay when it's fixed, but journalists I think it's very important to also remember to tell the reader "Okay, we had a mistake here". For example now in our automated Covid graphs we have, like, this kind of long list of notes (...), they are hidden, but they are in one box and then you open it and then there's all kinds of notes telling "Okay, correction: we had a mistake there, now it's fixed and it's like this".
>
> (Computational journalist, *Helsingin Sanomat*, Finland)

Traditional ways of organising journalistic work were also cause for miscommunication between editorial staff and technologists. The senior technologist at the Finnish broadcaster YLE shared that top-down approaches that usually prevail in media organisations made technologists feel a bit like they were the "IT department":

> In the old approach, our department would be more like [an] IT department [where] some journalists come and ask about: "Could you build some kind of tool for us?" And there would be some kind of miscommunication anyway in that and some kind of fear for each other, that we do not know enough about journalism or we do not know enough (…) about technology.
>
> (Senior technologist, YLE, Finland)

He said that problems like these were solved after the team managed to have a form of Agile workflow in place to be able to work with them, which entails a more horizontal way of working where journalists have to sit with technologists to see their projects through. "We are working together, not like (…) you come and order something from the IT department like [a] new computer, which was the old world [where] basically every nerd was someone to be afraid of", said the senior technologist. To act as a liaison with the rest of the newsroom, a journalist who is well-versed in computational matters was hired on the team, where he also writes templates for automated news and answers the team's questions on journalistic issues.

6.3.2   Limitations posed by platforms

As for tensions between newsrooms and players outside the journalistic field—in this case Big Tech companies—a certain type of struggle was visible, to a degree, in news organisations' efforts to make automated news content correspond to search engine criteria. This was in fact one of the reasons why the Spanish digital newspaper *El Confidencial* stopped its collaboration with Narrativa:

> There's, like, some positive conclusions about this project, but also negative [ones]. Some of them [were] the articles—the automatic articles—they are not optimised for SEO (i.e., search engine optimisation), they don't have tags. So there's a lot of things to polish in order to implement it in a production environment correctly. So then (…) we decided to create our own solution inside the company.
>
> (Executive, *El Confidencial*, Spain)

Having their own automated solution instead enabled *El Confidencial* to have more flexibility with regard to including results from headline A/B testing, a computational process whereby different headlines are shown to readers online and which determines those that work best (see Hagar, Diakopoulos & DeWilde, 2022). Yet, even while trying to conform to SEO criteria, the executive at the Spanish broadcaster RTVE reminded the need to remain distinct from other organisations:

> The algorithm that they [Google] have or they are using probably will understand that our news is very related to others. So we try to break this, changing the language and trying to [have] specific words for example, or changing the headline or the leads, to change the style. So probably the origin of the news is the same that they have [in] other newspapers or webpages, but there is the final result: we try to be different from the others.
>
> (Executive, RTVE, Spain)

In parallel with search engine considerations, matching the format required to have automated audio news featured on some of the platforms' voice assistants also involved having to adapt to their specifics. As an example of this, the ABC manager pointed out that a lot of work was needed to get Amazon's virtual assistant to accurately recognise all the towns and suburbs that they wanted to include in their automated emergency summaries (see Fell, 2021b).

In this last section, I have tried to cast light on the types of tensions that occur between editorial staff and, on the one hand, newsroom technologists and, on the other hand, platform companies. Although those were mostly covert, tensions between editorial staff and technologists could be observed in the way technologists

expressed their frustration towards editorial staff's handling of abstraction and numbers, whereas editorial staff rather positioned themselves as expert sources on journalistic matters and routines. As for tensions between newsrooms and platform companies, a certain type of struggle could be seen in the way news organisations modify automated news content so as to match search engine criteria. In the next and final chapter, I will summarise my empirical findings and see how they connect to my research questions.

7 DISCUSSION AND CONCLUSION

In this conclusive chapter, I will provide a summary of my empirical findings, first by answering my research questions, then by reflecting on their main takeaways. I will also provide guidance as to practical applications of these findings as well as recommendations for future research. Finally, I will touch on the limitations I have faced in this project.

7.1 ANSWERING THE RESEARCH QUESTIONS

In this section, I will come back to my empirical findings as I delve into the three research questions I have set out in section 2.4. First, I will answer RQ1 by detailing media practitioners' perceived impacts of automated news on the work they do. Then, I will reflect on the considerations these entail for journalism practice and for journalism as a whole, thus addressing RQ2. Lastly, I will ponder on how these reflections contribute to our understanding of the relationship between journalism and technology (i.e., RQ3).

7.1.1 Perceived impacts of automated news

To answer RQ1 (i.e., *What are media practitioners' perceived impacts of automated news on the work they do?*), I will refer to the micro, mezzo and macro perspectives I detailed in section 4.1: at the micro level, I will list practical impacts of automated journalism; at the mezzo level, I will touch on organisational impacts; and at the macro level, I will address larger impacts as I look at wider forces or mechanisms that are behind both practical and organisational impacts. At the micro level, first, some of the practical impacts of using automated news can be found in the way media practitioners dealt with external datasets. As shown in section 6.1.1, their work was hindered by data releasing strategies (i.e., having to deal with multiple levels of government, uneven times of release, sudden changes of format) and data quality concerns (i.e., missing context, null or missing data). These challenges prompted

newsrooms to come up with their own computational solutions (see section 6.1.2), such as computational safeguards to verify automated news being produced at scale, shared spreadsheet systems in which journalists can input new COVID-19 numbers for the location they are based in, or algorithmic alerts that let journalists know when new data is out or warn them of any sudden changes of format.

As for the more organisational impacts of automated news, these were rather visible at a mezzo level as they concern procedures like the data quality control step implemented at the BBC (i.e., "building the story model"), where editorial staff were asked to exert critical thinking when including data into automated news templates (see section 6.1.2). Organisational impacts could also be seen in the discussions that were held around how to embed journalistic standards and practices into code for automated news (see section 6.2.2). Finally, introducing new forms of workflows like Agile-inspired methods—a horizontal way of working that departs from more traditional top-down approaches to journalistic work—could be counted as organisational impacts as well (see section 6.3.1).

Looking at the macro level now, one of the wider forces that may help explain the impacts of automated news has on the work of media practitioners relates to their ability to envision journalism both as a one-off endeavour—or, to put it in the words of the BBC editor, as a "beautiful piece of work that's completely owned by you and [has] nothing to do with anybody else"—and as a process that can be deconstructed in an abstract way close to computer programming (see Dierickx, 2021). As illustrated in section 6.2.1, this type of heuristic is at play in the application of structured journalism—a process of "atomizing the news" to transform narratives into databases (Caswell & Dörr, 2018; Anderson, 2018; Jones & Jones, 2019)—to envision, first, what an ideal story will look like, and then to break it down into smaller elements that can be reusable across many versions of that same story. That being said, envisioning journalism both as a one-off endeavour and as an abstract procedure similar to computer programming seems to be creating a gap between news workers who are able to comprehend this new type of reasoning and those having a harder time coming to grips with it.

7.1.2   Implications for practice and for journalism as a whole

To answer RQ2 this time (i.e., *What do these considerations entail for journalism practice and for journalism as a whole?*), I will evaluate the repercussions these perceived impacts have on journalism practice and on journalism as a whole by resorting to the critical lenses brought by Bourdieu's Field theory and ANT. From a Field theory perspective, first, the wider mechanisms that have to do with considering journalism both as a one-off endeavour and as a process that can be deconstructed in an abstract way close to computer programming point out to a new type of cultural capital (journalistic *and* technological) that media practitioners need to acquire, which I will call here *distinct-abstract* capital. Possessing this type of capital can translate into an easiness to engage with new technology-oriented forms of computational journalism that are progressively gaining traction within newsrooms (see section 2.2.3), thereby creating a new form of *news habitus* (Schultz, 2007). This form of *habitus* is, in turn, most likely to be picked up among new entrants like computational journalists, who best know how to mingle the specifics of journalism practice with abstraction concepts brought by computer programming. That being said, the growing emphasis that is put on acquiring this new type of *habitus*—as exposed in new recruiting strategies in section 6.2.1—could also result in creating a situation of *hysteresis* or a "Don Quixote effect", whereby practitioners who acquired their dispositions using a more traditional form of journalistic capital (e.g., storytelling, finding "exclusives", etc.) and who are unable to adapt to this new context may consequently be lagging behind. To a certain extent, this could be perceptible in the some of the tensions that are occurring within the journalistic field (see section 6.3.1), where technologists expressed their frustrations towards editorial staff's handling of abstraction and numbers while media practitioners positioned themselves as expert sources on journalistic matters and routines.

Looking now at whether these computational journalists—or teams of journalists working with technologists—contribute to either changing or reinforcing the journalistic *doxa* when working with automated news, organisational impacts like implementing a data quality control step at the BBC, embedding journalistic standards and practices into code and resorting to Agile-inspired methods to work with technologists all point out to the influence that the technological field is having on the journalistic one. We can therefore posit that the deployment of automated news within newsrooms implies significant changes to journalism practice, thus

contributing to modifying the prevailing *doxa* in the field. At the same time, it is also worth reflecting on whether technologists' own *doxa* complements Deuze's (2005) *ideal-typical values* of *public service, objectivity, autonomy, immediacy* and *ethics*, which I argued are representative of the journalistic one. Wu, Tandoc and Salmon (2019c) describe the technological *doxa* that surrounds algorithmic automation as being rooted in a "Silicon Valley ethos" that places a high value on open data, client feedback and collaboration with other technological firms, as well as on digital literacy and awareness: there are then obvious connections to be made with computational journalists' ideals of public access to information and collaboration outside the newsroom (see section 2.2.3), which constitute in themselves a departure from standard journalism practice where "exclusives" are highly sought-after and where journalists generally act as gatekeepers in news selection. Even though Wu, Tandoc and Salmon indicated that there may be commonalities between the technological and journalistic *doxas*—most notably around audience needs—these are nonetheless clear evidence of technological *heteronomy* within the journalistic field, which may become even more pervasive as the authors stress that the digital turn within newsrooms could eventually result in technologically-minded agents becoming more dominant in the field.

Aside from this technological-bound form of *heteronomy* and from obvious cost-saving considerations occasioned by the use of automated news (see Kim & Kim, 2017), more traditional manifestations of economic *heteronomy* driven, this time, by audience and marketing needs could still be observed, as when news organisations strive to adapt automated news to SEO requirements or platforms' technical standards (see section 6.3.2). This is in line with the digital *heteronomy* identified in Lindblom, Lindell and Gidlund (2022), which brings a *virality* type of capital and an engagement form of *habitus* to light. Besides, conforming to SEO requirements also speaks to Bourdieu's *de-differentiation* caveat on commercial homogeneity within the media (see section 3.2.2): in a sense, this could be interpreted as a continuation of the shift from political *heteronomy* in the journalism landscape to audience, advertising and, now, techno-commercial forms of *heteronomy* (see section 2.2).

Using ANT's lenses now, considering automated news as *mediator* pointed out to new meaningful *translations* that are visible through the enrolment of alternate data

sources (i.e., a media organisation's own internal feed like the BBC's and ABC's election results feed, own data collection like Reuters' and AFP's manual input of COVID-19 statistics, or archives as in STT's use of past sports reports to train machine learning models—as well as crowdsourced material like *Stuttgarter Zeitung*'s use of community sensor data or, else, social media feeds as in RTVE's automated football stories that reflect readers' own preferences), the enticement of journalists while putting them at the centre of interfaces that are designed internally like self-editing tools (e.g., Reuters, El Confidencial), notification streams (e.g., BBC, Reuters) and/or search features to access automated backgrounders (e.g., *The Washington Post*) and, finally, the enlistment of vocal elements as in NLG-to-audio output (e.g., ABC, *The Washington Post*). This makes the movement of what can be considered the "actor-network of automated journalism" discernible to the researcher's eye and, hence, gives an indication as to where it is heading. All in all, this testifies of a growing journalistic professionalisation in the way automated news is being employed, as it is drifting away from political and commercial influences (i.e., public service data, data brokers, automated content providers and third-party self-editing tools) to become more under journalists' control, but also in citizens' hand (i.e., using crowdsourced material as a source). As shown in Table 6, North Atlantic media organisations (i.e., Reuters, *The Washington Post*, ABC, BBC) clearly lead the way in this process of *differentiation*, in accordance with Hallin and Mancini's (2004) typology (see section 3.2.2): as they write (*ibid.*, p. 80), "the Liberal Model is characterized by a high degree of differentiation of the media from other "other social bodies," particularly those historically active in the political sphere", which in this case also applies to techno-commercial influences. Hence, BBC's and ABC's use of internal feeds, Reuters' own in-house self-editing tool and the *Washington Post*'s providing access to automated backgrounders—to name a few—all contribute to greater journalistic professionalisation by ensuring independence from all these forms of external influences.

That being said, a process of *de-differentiation* could also be at play in that compliance with platforms' terms and conditions is generally needed to be able to connect to social media APIs (see van Dijck, Poell & de Waal, 2018) and matching their technical standards is necessary to have automated audio stories featured on voice assistants (e.g., Amazon's Alexa, Google Assistant). To a degree—and in line

with Prior's (2008) *prism* metaphor where ANT and Field theory meet—this goes back to Bourdieu's argument on commercial homogeneity within the media which, according to him, carries the risk of bringing "uniformity, censorship and even conservatism" (2005a, p. 44) among news media. The question as to whether platforms or news organisations will act as *spokespersons* in this growing actor-network of automated journalism then remains open: should news media take on this role, for instance while developing their own self-editing solutions or relying on internal feeds, this could be interpreted as reinforcing the *autonomous* pole of the journalistic field, whereas—should they become too dependent on Big Tech companies for data acquisition and dissemination of automated news products—this may result in making the field even more porous to techno-commercial *heteronomy*.

7.1.3   Better understanding the relationship between journalism and technology

Touching on RQ3 now (i.e., *How can these reflections advance our understanding of the relationship between journalism and technology?*), I will reflect on how some of the dimensions I touched on in this thesis may contribute to our understanding of the relationship between journalism and technology. First, I will look at Schudson's (1978) idea of a mutual shaping relationship between journalism and technology; second, at Örnebring's (2010) argument about journalism practice being shaped by the influence that societal context holds on technology use; and, finally, at Powers' (2012) three categories of how media practitioners react to new technological capabilities being brought into the newsroom (see section 1.1.1). With regard to Schudson's account of a mutual shaping relationship between journalism and technology, it seems that both datafication and the new "Spring of AI" (see section 2.1) are influencing the development of automated news products: on the one hand, increased datafication led to an even higher number of open data being released (e.g., open government, crowdsourced content) while, on the other hand, machine learning started making its way into NLG production (e.g., STT's use of archival material to generate automated news). That being said, contrarily to Schudson's observations on American newspapers driving the demand for new means of production (i.e., the steam-powered printing press), the news media do not constitute, this time, a significant demand that contributes to NLG development: in

fact, they only represent a slim share of clients advertised on NLG companies' websites, which feature more prominent sectors of activity like real estate, retail or government services[33]. This is in line with Sirén-Heikel, Kjellman and Lindén's (2022) remarks that NLG companies improve their products using journalistic expertise, but do so to go after more lucrative industries (see also Wu, Tandoc & Salmon, 2019c).

Looking at Örnebring's point about journalism practice being shaped by the influence that societal context holds on technology use, my study highlights a shift in recruiting strategies, which gives an edge to editorial staff who are familiar with abstraction and numbers or technologists having a good grasp of journalistic issues (see section 6.2.1). To some extent, we may wonder whether this emphasis is to be linked with the same mass data collection and processing context that has given rise to *surveillance capitalism* (Zuboff, 2019), where personal data is being processed for commercial purposes. Therefore, just like the capitalistic logic of competition saw using technology to increase production—resulting in an *ideal of speed* in journalism—new forms of competitiveness where data science and algorithms are employed for mass data collection and processing could bring about a new *computational ideal* within newsrooms.

As per Power's evaluation of how news staff react to new technical capabilities being brought into the newsroom, I have made the case above for automated journalism not to be considered under his category of a mere extension of existing occupational practices and values, as it brings forth organisational impacts like inspecting data quality prior to writing templates, embedding journalistic standards and practices into code and adopting more horizontal workflows like Agile-based methods, all of which point to the influence that the technological field exerts on the field of journalism. As it stands, the introduction of automated news within newsrooms rather corresponds to Powers' category of a technology-inspired form of work that can be used as a basis for a reinvention of current occupational norms, even though what lies ahead is still unclear. As for seeing new technological capabilities as a threat that needs to be subordinated because they do not correspond to occupational norms—which constitutes Powers' last category—it may be worth

---

[33] These were verified on the websites of some of the main NLG content or platform providers (Arria, no date; Automated Insights, no date; Syllabs, no date).

pondering on whether it is the usual depiction of a robot journalist stealing journalistic jobs that matters the most (see Lindén & Dierickx, 2019), as opposed to journalists not benefiting from a distinct-abstract kind of capital, or to put it differently not being able to conjugate the specifics of journalism practice with the level of abstraction required in computer programming.

7.2   CONCLUSIVE REMARKS AND FURTHER REFLECTIONS

I will conclude here by reflecting on the key takeaways that these research questions brought and point out to research that can help further advance them. I will also specify the degree to which my PhD dissertation contributes to operationalising theory and to the field of journalism studies, and how it can be used for theoretical generalisation. Finally, I will list out practical recommendations for more industry-focused uses of this research thesis and explore the extended labour impacts of automated news, before finishing off on limitations and future research.

7.2.1   Research thesis' key considerations

My research has shown that, like any other computational journalism projects, the implementation of automated news translate into giving an edge to media practitioners who are able to conceive journalistic work both as a one-off endeavour and as a process that can deconstructed in a an abstract way close to computer programming. This is, by far, the most central element to my research thesis as it made me come up with this whole new concept of a distinct-abstract capital, which falls under Bourdieu's understanding of a cultural capital (i.e., unique abilities to a field) that regroups both technological and journalistic capital (see Bourdieu, 2005b and Schultz, 2007). In practice, possessing it translates into giving a head start to editorial staff already equipped with at least a basic understanding of data science and algorithms, which is certainly of help when co-creating automated news using self-editing tools or when working directly with it in a form of combined workflow. Having said that, possessing these types of prerequisites would also allow them to program automated news on their own and set up new iterations the likes of alert systems and computational safeguards. Participating in the development and

implementation of automated news this way then further legitimises media practitioners in having a say over the collective impacts that it implies, like rethinking editorial procedures (e.g., embedding journalistic standards and practices into code). In sum, while computational journalists are expected to be a natural fit for this type of profile, the question as to whether shoe leather and other rank and file journalists will be able to adapt remains widely open. This may turn out to be a central issue as other forms of computational journalism like data journalism and using data mining techniques in investigative journalism could become more prevalent in the years to come, and even perhaps dominating features as discourses around a "computational ideal" within newsrooms gradually make their way through. As for developing internal automated journalism software, there seems to be a push towards greater journalistic professionalisation as media organisations rely on datasets other than private and public service ones, and as journalists participate in ways other than through the affordances already provided for by external self-editing tools. That being said, adapting to platforms' requirements like including SEO-friendly keywords and complying with their technical standards to have, for instance, automated audio stories featured on voice assistants could also be seen as a hurdle to journalistic autonomy.

To move past some of these challenges, guidance can be found in scholarship and policy papers written on automated news or on news innovation in general. In terms of organisational frameworks, Mills and Wagemans (2021) described how in-house media labs—including some that are documented in my study (e.g., BBC News Labs)—break from traditional Research and Development departments as they typically rely on design thinking and Agile-inspired methods so as to be able to implement change at speed. They also mention that these can be conceived as "self-contained units" that are governed under different rules than those of the main newsroom, yet reflect the organisation's strategic goals. Besides, Cools, Van Gorp and Opgenhaffen (2022, p. 13) remarked that "the higher the integration with the larger newsroom, the merrier these news lab members are likely to consider themselves as journalists or news workers", which connects to Diakopoulos' point (2019) on embedding computational thinkers in a newsroom environment (see section 2.2.3).

With regard to setting up automated news, Dierickx (2020) suggested using ISO 9000 principles—an international set of references in quality management—in order to satisfy needs when designing this type of product: this involved for instance journalists' participation in programming a newsroom bot as they were considered end users in this case. When it comes to more advanced machine learning models, news organisations' own set of standards like the BBC's Machine Learning Engine Principles (Macgregor, 2021), the Bavarian broadcaster Bayerischer Rundfunk's AI Ethics Guidelines (Bedford-Strohm, Köppen & Schneider, 2020) or the Norwegian media group Schibsted's FAST framework (Størmer Thaulow, Næss & Stenbom, 2021) can be seen as potential guidance for an "Ethics By Design" approach to automated news, as they essentially address fairness, accountability, confidentiality, transparency and safety concerns in information and communications technology (see Olteanu *et al.*, 2019), which could furthermore match the prerequisites of regulatory frameworks like the draft European Union Artificial Intelligence Act (see Helberger & Diakopoulos, 2022).

Additionally, Happenen (2020) raised that media councils should also play a role in setting out standards for the implementation of automated news, for instance when a complaint is lodged against an organisation's use of algorithmic automation and personalisation. The risk, he wrote, is that other bodies like national legislators, the European Union or platform companies may otherwise be filling that gap, thus potentially jeopardising press freedom. As for negotiating with platforms in the process of complying with some of their requirements (e.g., SEO, technical standards), Lindén *et al.* (2022) reported that adopting common data management standards across media organisations could further reinforce their independence *vis-à-vis* Big Tech companies.

7.2.2   Contributions to journalism studies and theory

This thesis contributes to operationalising theory through wielding Field theory to look into the more critical and interpretative aspects of using automated news. This theoretical operationalisation is made through a series of reflections around Bourdieu's idea of new entrants in a field, which are either changing or reinforcing the prevailing *doxa*. To ponder on this with respect to automated journalism, I

evaluated each of Wu, Tandoc and Salmon's (2019b) three areas of tensions as identified in their *Field analysis of journalism in the automation age* (i.e., structures external to the journalistic field, accumulating cultural capital, adversarial reactions within the journalistic field) against Deuze's (2005) five *ideal-typical values* of journalism's ideology (i.e., *public service, objectivity, autonomy, immediacy* and *ethics*), which I argue are representative of the journalistic *doxa*. This helped me delineate three themes that are relevant to examine media practitioners' reactions to automated news: first, the over-reliance on external datasets (i.e., structures external to the field); second, the need for them to adopt a computational thinking mindset (i.e., accumulating cultural capital); and, third, conflicts (i.e., adversarial reactions within the field) between editorial staff and, on the one hand, businesspeople and technologists in news and, on the other hand, players external to journalism, in this case Big Tech companies. These reflections were summarised into Table 3 so that they could be readily available to develop questionnaires for my interviewees. I also relied on these themes in the coding procedure that I set out for my data analysis, and used them as a guiding thread in chapter 6. To a certain extent, it is worth considering whether this way of operationalising theory could be used to examine other algorithmic-driven news products and practices, like automated fact-checking and employing data mining techniques in investigative journalism (see section 1.2.1).

In parallel, this dissertation has filled an important research gap in journalism studies since it documents the practical, organisational and more deeper impacts of automated news on the work of media practitioners, using both Bourdieu's Field theory and ANT to look at this with a critical eye. Field theory is indeed especially well-suited to unveiling internal tensions that pertain to the deployment of automated news as it helps account for both structure and agency. That being said, ANT brings valuable insights too, most notably in relation to how automated news is being transformed as it is deployed within newsrooms. All in all, this thesis constitutes a substantial contribution to practice-oriented scholarship on automated journalism, as much because of the way theoretical frameworks are being used as for the sheer number and geographical distribution of news organisations included in this study (i.e., 23 organisations in 13 countries), this together with exclusive insights on how automated news were used during COVID-19 and at the BBC.

At last, my strong focus on using Field theory enables me to make a claim to theoretical generalisation based on the concept of a distinct-abstract capital which, again, constitutes the very key finding of this PhD thesis: indeed, the ability to comprehend a subject—here journalism—both as a one-off endeavour and as a process that can be deconstructed in an abstract way close to computer programming could in fact maybe be observed in other fields of cultural production. In the journalistic field, I have shown that resorting to this type of distinct-abstract capital may contribute to redefining the *news habitus* (Schultz, 2007) as we have known it, as data science and programming skills seem to have become as important as making use of a more traditional journalistic kind of capital, like resorting to one's journalistic "gut feeling" (*ibid.*) or finding "exclusives" to work on, just to name a few. Such a transformed *habitus* can result in modifying rather than preserving the prevailing *doxa*, as made evident in the way computational journalists or teams of journalists working with technologists bring about important changes to journalism practice (i.e., new workflows, organisational procedures) as they implement automated news or other forms of computational journalism projects, thus showing a form of technological *heteronomy*. In this process, individuals who are unable to adapt their existing *habitus* to these changing circumstances—most probably because they are uncomfortable dealing with abstraction and numbers—could be facing a situation of *hysteresis* where they risk lagging behind. This could concern shoe leather or other rank and file journalists who are very well-versed in the specifics of journalism practice (e.g., storytelling, finding "exclusives", liaising with sources), but have a harder time coming to grips with programming essentials. To a certain extent then, could the same dynamics be observed in other fields of cultural production? For instance, in the field of arts, could it be that artists engaging with computer-generated content (i.e., "generative art") contribute to modifying the artistic *doxa* with a new *habitus* that is based on the same type of distinct-abstract capital?

7.2.3 Recommendations for practitioners and larger labour impacts

This PhD research has highlighted some of the main automated news features used during COVID-19, at the BBC and, more largely, across media organisations.

Consequently, I will share here some best practice guidelines and recommendations that are based on my own observations as well as others'. These go as follows:

- For large organisations that have technologists working with journalists on news innovation projects like automated journalism, careful thought should go into not considering technologists simply as an extension of the "IT desk". Setting up structures such as media labs that operate under different rules than those of the main newsroom, yet reflect the organisation's strategic goals (see Mills & Wagemans, 2021), could facilitate this process, in addition to including the right amount of Agile-inspired methods into traditional ways of organising journalistic work. To determine an appropriate framework to work with, design thinking workshops with both editorial staff and technologists could be organised so as to collect their insights. Additionally, proper intellectual property arrangements between employers and employees should be made ahead of time, for instance in employment contracts.

- To be able to participate in the design phase of automated news using self-editing tools (e.g., Arria NLG Studio, Reuters' Lynx Insight) or to work directly with it in a form of combined workflow (e.g., BBC's coverage of the 2019 general election in the United Kingdom), news workers need to be familiar with the type of abstract reasoning that is at the core of computer programming, in addition to their regular journalistic skills. Consequently:

  - Outsourcing automated news to an external content provider, having journalists co-creating it through self-editing tools or developing automated news in-house could be evaluated against news workers' overall level of comfort to work with abstraction concepts. As BBC News Labs did as part of its recruiting strategy, a small abstract reasoning test could help determine the right strategy to go for, and therefore help foresee budgeting costs. Indeed, resorting to a third-party tool could be priced at between 300 and several thousand euros a month (AX Semantics, no date 2) and outsourcing automated news to an external content provider could cost between 10,000 and 50,000

dollars a month (Mullin, 2015). In-house solutions would depend on salary scales for computational journalists or technologists able to engage with these types of products.

○ Following Diakopoulos (2019), training "tech-savvy" computational journalists who are best able to conjugate the type of abstract reasoning that is at the core of computer programming with the specifics of journalism practice could be done either: by embedding computational thinkers in an editorial environment; by training journalists in statistical methods, which acts as a prerequisite for advanced computer-assisted reporting (see Meyer, 1973); through revamped journalism curricula that would include elements of or be focused on acquiring a computational journalism knowledge.

● At an organisational level, the use of automated news—just like any other computational journalism projects—calls for important changes that have to do with having to rethink journalistic rules and routines. This could be as much about new editorial procedures (i.e., data quality control step, embedding journalistic standards and practices into code) as it is about new ways of organising journalistic work (i.e., adding Agile-inspired methods to a newsroom's workflow). This could be facilitated through:

○ Applying ISO 9000 principles— an international set of references in quality management —so as to satisfy needs when designing this type of product (Dierickx, 2020), or developing a set of "computational journalism guidelines" to translate a news organisation's own policies into code for automated news.

○ For advanced machine learning models, following one of the frameworks developed by media organisations like the BBC's Machine Learning Engine Principles, the Bavarian broadcaster Bayerischer Rundfunk's AI Ethics Guidelines or the Norwegian media group Schibsted's FAST framework. These can provide guidance for an "Ethics

By Design" approach to automated journalism, as they essentially address fairness, accountability, confidentiality, transparency and safety concerns in information and communications technology (see Olteanu *et al.*, 2019), which could help match the prerequisites of regulatory frameworks like the draft European Union Artificial Intelligence Act (see Helberger & Diakopoulos, 2022).

- With regard to in-house development of automated news systems, special attention should be given to: bypassing public service or private data, for instance while relying on a media organisation's own internal feed, data collection or archives as well as on crowdsourced content or social media feeds; using systems that involve journalists' participation in ways other than through the affordances already provided for by third-party tools (e.g., in-house self-editing tools, automated notification streams, automated backgrounders); generating output other than text, for instance NLG-to-audio summaries that can be inserted in a podcast or be accessed *via* a platform's voice assistant (e.g., Amazon's Alexa, Google Assistant). However, careful thought should go into finding the appropriate balance between complying with platforms' requirements (e.g., SEO, fitting technical standards) and maintaining journalistic autonomy. Common data management standards across media organisations (Lindén *et al.*, 2022) can be seen as a possible avenue.

On the more societal impacts of automated news, prior scholarship (see section 2.3.2) have discussed whether this type of technology could affect legal matters such as defamation lawsuits or copyrights claims (see Lewis, Sanders & Carmody, 2019; Weeks, 2014; Díaz-Noci, 2020) or business aspects like automated news' market effects (see Blankespoor, deHaan & Zhu, 2018). That said, automated journalism is much more likely to have a greater effect on civic life, as affordances like being able to tailor it to one's location or personal preferences (e.g., summarising the outcome of a sports game in a way that would read differently based on whether readers support the team or not) could further reinforce "filter bubbles" (Pariser, 2011): it would then be liable for making news readers see the world in a way that reflects their

own set of views (Graefe, 2016; Kim & Lee, 2019; Jia & Johnson, 2021), in addition to leaving journalists' gatekeeping role to algorithmic systems often criticised for their opacity (see Gillespie, 2014).

As for the larger labour impacts of automated news, my research has made clear that a reinvention of journalistic standards and practices that would take computational journalism aspects into account is needed. This contrasts with the usual debate over whether automated journalism "steal" journalistic jobs or, on the contrary, alleviates journalists' workload (see van Dalen, 2012; Carlson, 2015; Lindén & Dierickx, 2019). Albeit these questions may have been important in initial responses to automated news, my research demonstrates that the thick of the matter now lies with media practitioners being able to adopt a "computational thinking" mindset (see Wing, 2008): if this is the case, not only will they be able to set up their own automated stories using coding skills or self-editing tools— they would also be able to engage with all of the other nitty-gritty aspects of computational journalism, such as web scraping in data journalism or using open source machine learning libraries (e.g., PyTorch, TensorFlow) as part of data mining efforts in investigative journalism. In addition, possessing this computational knowledge would equip them with an in-depth understanding of how algorithms operate, which can be a critical skill to have when engaging with algorithmic accountability reporting (see Diakopoulos, 2015). In short, news workers who "think computationally" will be able to keep a close eye on questionable uses of algorithms for surveillance purposes (e.g., unlawful or unethical data collection for security or commercial goals) or in a way that hurts vulnerable groups (e.g., algorithmic biases in credit score or recruiting policies), as discussed in section 2.1.3.

7.2.4   Limitations and future research

As mentioned in my methodology, a first limitation to this research relates to the impossibility of carrying out direct observations because of the COVID-19 pandemic. Although newsroom ethnography was initially considered—and even arranged for with a couple of newsrooms—these plans had to be cancelled when it became evident that the pandemic would last for longer than initially envisioned. Instead, I made use of remote semi-structured interviews and of elements of a netnography, and

increased the number of news media under study. In consequence, I was not able to see how automated news was being used with my own eyes (although a virtual walkthrough was conducted with BBC): this resulted in me walking a fine line between findings I could directly document, like an automated news dashboard that is available online, and others that were reported or not directly visible to me, like details of a computer script. Even though I did my best to verify all of these elements, they are still exposed to the type of fallibility that goes with human interpretation (see Benton, 2004; Clark, 2008).

Moreover, not being able to conduct fieldwork on the ground made it more difficult to document potential tensions within newsrooms, which resulted in section 6.3 being significantly shorter than the other ones. This also prevented me from examining whether automated journalism is indeed being used to alleviate the work of media practitioners so that they can focus on more in-depth reporting or, on the contrary, is being leveraged so that journalists are assigned to stories that reflect the priorities of the marketing department instead (e.g., "clickbait" stories, native advertising). On this latter point, not being able to interview the marketing and advertising teams as part of this research project (see section 6.3) can be seen as another hindrance, especially as their opinions on including SEO keywords and complying with platforms' requirements when setting up automated news would be worth having.

A second limitation to this study has to do with a very much Western-centric selection of media organisations: at the time I reached out to interviewees, automated news was still a relatively new development that seemed to concern mostly news organisations based in the West, as well as some Asian newsrooms that could not efficiently research because of my own language limitations (see section 2.3.1). This meant I could not document the use of automated news in certain regions, like South America or East Asia. That being said, an growing number of scholars are now looking into these areas, among which figure research on the way automated news is employed at the Czech news agency ČTK (Moravec *et al.*, 2020) and across South American news media (García-Perdomo, Montaña-Niño & Magana, 2022).

Finally, one last limitation relates to not being able to set in stone what remains essentially a field in flux, where new technical breakthroughs or ways of implementing automated journalism could be happening as I am writing these lines.

For example, at the start of my PhD, most NLG companies appeared to be external content providers only, in charge of creating automated news products in place of media companies: it turned out later that they also started offering self-editing tools as well, as in the case of Automated Insights (see Mullin, 2015). This fast-paced evolution of automated news products makes it difficult to analyse them based on development types (i.e., external content providers, in-house, third-party self-editing tools); however, this could be done once this is stabilised enough.

Besides, this research still remains exploratory in essence, as it addresses characteristics of automated news that concern, after all, a limited number of media organisations (i.e., 23 of them, based in 13 countries). To verify some of this thesis' conclusions and strengthen their claims to generalisation, a larger sample of news organisations could be envisaged to conduct, for instance, an survey among news workers.

In terms of future research—outside the need to evaluate the deployment of automated news beyond Western countries and readers' perceptions of more advanced machine learning models (see Danzon-Chambaud, 2021a)—special attention should go into analysing how discourses around automated news alleviating or augmenting the work of media practitioners fare against a potential situation of *hysteresis* experienced by those unable to acquire a distinct-abstract kind of capital (e.g., shoe leather and other rank and file journalists). This could be investigated using Meyer and Rowan's (1977) notion of *coupling* and *decoupling* in New Institutionalism research, where an organisation's symbolic practices, or *rational myth*, are either connected (i.e., *coupling*) or disconnected (i.e., decoupling) to its day-to-day operations, or *technical core*. This would help answering questions such as: Do these discourses that seem to reflect a news organisation's *rational myth* translate into actual newsrooms practices (i.e., *coupling*)? Or, on the contrary, are these separated from the organisation's own *technical core* (i.e., *decoupling*)?

Other than this, possible research avenues include using ANT to determine whether, in the ongoing assemblage of an "automated journalism actor-network", news organisations or platforms act as *spokespersons*, especially as it may turn into a *macro actor* able to restructure media production as a whole. To a certain extent, platforms can be seen as already gaining the upper hand as recent text summarisation

efforts—which are somehow related to automated news—appear to be quite tailored to fitting social media content (see section 1.2.2). Such an analysis would be essential in determining power relationships likely to shape future developments of automated news products. As for the current direction that the actor-network of automated news is taking, it would be worth evaluating whether systems that call for journalistic participation—other than through the affordances already provided for by third-party self-editing tools—reflect, again, a media company's *rational myth* of increased efficiency (i.e., *coupling*) or help mitigate potential *hysteresis* among news staff, while giving them the *illusio* that it is worth being involved in content creation this way. To complement all these, a political economy analysis of the main funders of automated journalism projects—whether these are universities, platform companies, foundations or governmental structures (see also Ferrucci & Eldridge II, 2022)—would be worth undertaking.

APPENDIX A. Corpus of empirical scholarship on automated journalism used in the systematic literature review, with coding procedures.

ID	V1_Reference	V2_Country	V3_First_Publication_Year	V4_Keywords	V5.1_Theory_Use	V5.2_Main_Theoretical _Background	V5.3_Theory_Specifics	V6_Method	V7_Investigation_Domain
1	Waddell, 2018	USA	2017	M	3	2	M	3	1
2	Caswell & Dörr, 2018	CHE-USA	2017	M	1	N/A	N/A	8	2
3	Haim & Graefe, 2017	DEU	2017	M	3	2	Expectation-Confirmation Theory	3	1
4	Linden, 2017a	FIN	2016	M	2	1	M	5	2
5	Kim & Lee, 2019	KOR	2018	N/A	1	N/A	N/A	M	1
6	Clerwall, 2014	SWE	2014	M	1	N/A	N/A	3	1
7	Young & Hermida, 2015	CAN	2014	M	1	N/A	N/A	M	2
8	Montal & Reich, 2017	ISR	2016	M	1	N/A	N/A	M	2
9	Liu & Wei, 2019	USA	2018	M	3	2	Expectancy Violation Theory	3	1
10	Dörr, 2016	CHE	2015	M	3	1	Institutional Theory	M	2
11	Lokot & Diakopoulos, 2016	USA	2015	M	1	N/A	N/A	2	2
12	van Dalen, 2012	DNK	2012	M	1	N/A	N/A	1	2
13	Carlson, 2015	USA	2014	M	2	4	M	1	2
14	Thurman, Dörr & Kunert, 2017	CHE-DEU	2017	M	1	N/A	N/A	M	2
15	Wölker & Powell, 2021	NLD	2018	M	1	N/A	N/A	3	1
16	Graefe, Haim, Haarmann & Brosius, 2018	DEU	2016	M	1	N/A	N/A	3	1
17	Lewis, Sanders & Carmody, 2019	USA	2018	M	1	N/A	N/A	1	1
18	Kim & Kim, 2017	KOR	2016	M	3	1	M	7	2
19	Kim & Kim, 2018	KOR	2017	M	3	3	M	M	2
20	Zheng, Zhong & Yang, 2018	CHN-USA	2018	M	3	3	M	3	1
21	Jung *et al.,* 2017	KOR	2017	M	3	3	In-group and out-group	3	3
22	Visvam Devadoss, Thirulokachander & Visvam Devadoss, 2019	IND	2018	M	1	N/A	N/A	8	2
23	Blankespoor, deHaan & Zhu, 2018	USA	2017	M	1	N/A	N/A	8	1
24	Melin *et al.*, 2018	FIN	2018	M	1	N/A	N/A	3	1
25	Jones & Jones, 2019	GBR	2019	M	1	N/A	N/A	M	2
26	Ford & Hutchinson, 2019	AUS	2019	M	1	N/A	N/A	M	3
27	Tandoc, Lim & Wu, 2020	SGP	2020	M	3	2	Expectancy Violation Theory	3	1
28	Wu, 2020	USA	2019	M	3	2	Cognitive Authority Theory	3	1
29	Waddell, 2019a	USA	2019	M	3	2	Expectancy Violation Theory	3	1
30	Waddell, 2019b	USA	2019	M	3	2	M	3	1
31	Túñez-Lopez, Toural-Bran & Valdiviezo-Abad, 2019	ESP	2019	M	1	N/A	N/A	1	2
32	Rojas Torrijos, 2019	ESP	2019	M	1	N/A	N/A	1	2
33	Díaz-Noci, 2020	ESP	2020	M	1	N/A	N/A	1	1

Multiple Theories	
ID	Theories
1	Expectancy Violation Theory
1	MAIN Model
4	Social Constructivism
4	Framing Theory
13	Technological Dramas
13	Grounded Theory
18	Institutional Entrepreneurship
18	Structural Inertia
18	Institutional Isomorphism
19	Innovation resistance theory
19	Institutionalism
20	High-context and low-context cultures
20	Holistic/analytic thinking framework
30	MAIN Model

Multiple Methods	
ID	Methods
5	3
5	5
7	1
7	2
7	5
8	1
8	5
10	5
10	8
14	2
14	5
19	5
19	8
25	5

Definitions	
M	Multiple results
N/A	Non applicable
NULL	No result
V5.1_Theory_Use	
No	1
Mere mention	2
At use	3
Other	4
V5.2_Main_Theoretical _Background	
Sociology	1
Psychology	2
Mixed	3
Other	4
V6_Method	
Content analysis	1
Ethnography/Observation	2
Experiments	3
Focus group	4
Interviews	5
Narrative analysis	6
Survey	7
Other	8
V7_Investigation_Domain	
Reach	1
Practice	2
Mixed	3
Other	4

180

APPENDIX B. Details of interviewees' news organisations, roles and gender as well as interview date and duration.[34]

Interview	Organisation	Country	Position	Gender	Date	Duration
1	RADAR	United Kingdom	Editor	M	30/06/2020	00:35:31
2	Canadian Press	Canada	Senior computational journalist	M	07/07/2020	00:44:16
3	Helsingin Sanomat	Finland	Computational journalist	F	14/08/2020	00:43:10
4	OMNI	Sweden	Manager	F	09/09/2020	00:33:09
5	The Times	United Kingdom	Computational journalist	M	10/09/2020	00:44:15
6	Stuttgarter Zeitung	Germany	Editor	M	11/09/2020	00:49:28
7	France Bleu	France	Manager	M	07/10/2020	00:31:29
8	YLE	Finland	Senior technologist	M	08/10/2020	00:48:30
9	AP	United States	Executive	F	15/10/2020	00:30:12
10	BBC	United Kingdom	Manager	M	26/10/2020	00:42:53
11	NTB	Norway	Editor; Executive	M; M	19/11/2010	00:37:43
12	Tamedia	Switzerland	Senior computational journalist	M	20/11/2010	00:44:27
13	Bayerischer Rundfunk	Germany	Senior technologist	M	23/11/2020	00:35:27
14	Washington Post	United States	Executive	M	30/11/2020	00:29:21
15	Bloomberg News	United States	Executive	F	03/12/2020	00:34:08
16	STT	Finland	Executive	F	04/12/2020	00:31:09
17	Rossel/Sudpresse	Belgium/France	Executive	M	09/12/2020	00:44:21
18	El Confidencial	Spain	Executive; Technologist	M; F	15/12/2020	00:34:05
19	RTVE	Spain	Executive	M	16/12/2020	00:49:56
20	ABC	Australia	Manager	M	22/12/2020	00:45:09
21	BBC	United Kingdom	Senior technologist	M	22/03/2021	00:23:14
22	BBC	United Kingdom	Computational journalist	M	24/03/2021	00:41:21
23	BBC	United Kingdom	Journalist	M	29/03/2021	00:32:03
24	BBC	United Kingdom	Assistant editor	F	01/04/2021	00:12:11
25	ANSA	Italy	Executive	M	01/04/2021	00:21:21
26	BBC	United Kingdom	Technologist (1)	M	06/04/2021	00:42:01
27	AFP	France	Manager; Senior journalist	M; M	07/04/2021	00:34:52
28	BBC	United Kingdom	Technologist (2)	F	16/04/2021	00:33:00
29	Reuters	United Kingdom	Editor	M	20/04/2021	00:44:38
30	BBC	United Kingdom	Editor	M	28/04/2021	00:28:34

---

[34] Positions are based on my own understanding of interviewees' roles and skills, and do not necessarily correspond to their official titles.

APPENDIX C. Questionnaire developed to interview an executive at the *Washington Post*.

- Could you please tell me a bit about yourself?

- I've been following what the *Post* has been doing with [the automated journalism software] Heliograf (Rio Olympics, 2016 elections), then I saw the Computational Journalism Lab for this year's elections. Could you please give me an overview of automated text generation projects at the *Post*?

- Why going for an in-house solution rather than outsourcing automated news to an external NLG provider or using a third party tool?

- How is editorial staff involved in creating automated news at the *Post*?

- When writing templates, how do they balance predicting elements of the story in advance with the uncertainty of news?

- How do you make sure data sources selected for automated news are diverse enough?

- What about important aspects of the story that cannot be automated?

- What safeguards do you have to prevent unwarranted content from creeping into the final copy?

- How do you make sure to embed the *Post*'s journalistic standards and practices into the conception of automated news?

- How did work routines/media coverage change at the *Post* after the implementation of automated news?

- What unforeseen situation did you run into while using automated news to cover this year's election?

REFERENCES

Ables, K. (2018) 'What happens when China's state-run media embraces AI?', *Columbia Journalism Review*, 21 June. Available at: https://www.cjr.org/analysis/china-xinhua-news-ai.php (Accessed: 8 September 2022).

Adams, K. (2013) 'Back to the Story Spine', *Aerogramme Writers' Studio*, 5 June. Available at: https://www.aerogrammestudio.com/2013/06/05/back-to-the-story-spine/ (Accessed: 8 September 2022).

Adams, K. (2022) '(Re)writing history: Examining the cultural work of the obituary and journalists' construction of a former president's legacy', *Journalism*, 23(5), pp. 1027–1043. Available at: https://doi.org/10.1177/1464884920929831.

Adler, P. (2008) 'Technological determinism', in S. R. Clegg and J. R. Bailey (eds) *The International Encyclopedia of Organization Studies: Vol. 4.* Thousand Oaks, CA: SAGE, pp. 1536–1539. Available at: https://doi.org/10.4135/9781412956246.n530.

Akremi, L. (2020) 'Generalisation', in P. Atkinson (ed.) *SAGE Research Methods Foundations*. Available at: https://dx.doi.org/10.4135/9781526421036931841.

Akrich, M. (1992) 'The de-scription of technical objects', in W.E. Bijker and J. Law (eds), *Shaping Technology/Building Society*. Cambridge, MA: MIT Press, pp. 205–24.

Allyn, B. (2022) 'Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn', *NPR*, March 16. Available at: https://text.npr.org/1087062648 (Accessed: 23 August 2022).

Altheide, D. (1976). *Creating Reality: How Television News Distorts Events*. Thousand Oaks, CA: SAGE.

Anderson, C. W. (2011) 'Between creative and quantified audiences: Web metrics and changing patterns of newswork in local US newsrooms', *Journalism*, 12(5), pp. 550–566. Available at: https://doi.org/10.1177/1464884911402451.

Anderson, C. W. (2013) 'Towards a sociology of computational and algorithmic journalism', *New Media and Society*, 15(7), pp. 1005–1021. Available at: https://doi.org/10.1177/1461444812465137.

Anderson, C. W. (2018). *Apostles of Certainty: Data Journalism and the Politics of Doubt*. Oxford: Oxford University Press.

Appelman A. and Sundar, S. S. (2016) 'Measuring message credibility: Construction and validation of an exclusive scale', *Journalism & Mass Communication Quarterly*, 93(1), pp. 59–79. Available at: https://doi.org/10.1177/1077699015606057.

Archer, L. (2016) 'Nine things we've learnt in one month on Facebook Messenger' [Medium], December 16. Available at: https://medium.com/abc-news-

australia/nine-things-weve-learnt-in-one-month-on-facebook-messenger-
f1c141ffe6ad (Accessed: 14 August 2022).

Arria, (no date). *Discover how Arria brings the power of language to your industry*. Available
at: https://www.arria.com/industry-expertise/ (Accessed: 20 May 2022).

Ausserhofer J., Gutounig R., Oppermann M., Matiasek, S. and Goldgruber, E. (2017) 'The
datafication of data journalism scholarship: Focal points, methods, and research
propositions for the investigation of data-intensive newswork', *Journalism*, 21(7), pp.
950–973. Available at: https://doi.org/10.1177/1464884917700667.

AX Semantics (no date 1), *AI-powered content generation solution for the digital age*.
Available at: https://en.ax-semantics.com/the-software/ (Accessed: 8 September
2022).

AX Semantics, (no date 2). *License Plans and Pricing*. Available at: https://en.ax-
semantics.com/pricing/ (Accessed: 20 May, 2022).

Ayres, L. (2008a) 'Semi-structured interview', in L. M. Given (ed.) *The SAGE Encyclopedia
of Qualitative Research Methods*. Thousand Oaks, CA: SAGE, pp. 811–812. Available
at: https://dx.doi.org/10.4135/9781412963909.n420.

Ayres, L. (2008b) 'Thematic coding and analysis', in L. M. Given (ed.) *The SAGE
Encyclopedia of Qualitative Research Methods*. Thousand Oaks, CA: SAGE, p. 868.
Available at: https://dx.doi.org/10.4135/9781412963909.n451.

Beasley, B. A. (2009) 'History of Journalism: Before 1861', in C. H. Sterling (ed.)
Encyclopedia of journalism: *Vol. 1*. Thousand Oaks, CA: SAGE, pp. 696–701. Available
at: https://doi.org/10.4135/9781412972048.n184.

Bedford-Strohm, J., Köppen, U. and Schneider, C. (2020) 'Our AI Ethics Guidelines',
Bayerischer Rundfunk, November 30. Available at: https://www.br.de/extra/ai-
automation-lab-english/ai-ethics100.html (Accessed: 09 August 2022).

Beer, D. (2017) 'The social power of algorithms', *Information, Communication & Society*,
20(1), pp. 1–13. Available at: https://doi.org/10.1080/1369118X.2016.1216147.

Beiler, M., Irmer, F. and Breda, A. (2020) 'Data journalism at German newspapers and
public broadcasters: a quantitative survey of structures, contents and perceptions',
Journalism Studies, 21(11), pp. 1571–1589. Available at:
https://doi.org/10.1080/1461670X.2020.1772855.

Belair-Gagnon, V. (2015) *Social Media at BBC News: The Re-Making of Crisis Reporting*.
Abingdon-on-Thames: Routledge.

Belair-Gagnon, V. and Holton, A. E. (2018). 'Boundary work, interloper media and analytics
In newsrooms', *Digital Journalism*, 6(4), pp. 492–508. Available at:
https://doi.org/10.1080/21670811.2018.1445001.

Bell, D. (1973) *The Coming of Post-Industrial Society: A Venture in Social Forecasting*. New York, NY: Basic Books.

Bell, E. J., Owen, T., Brown, P. D., Hauka, C. and Rashidian, N. (2017) *The Platform Press: How Silicon Valley Reengineered Journalism*. Tow Center for Digital Journalism report. Available at: https://doi.org/10.7916/D8R216ZZ.

Benson, R. and Neveu, E. (2005). 'Introduction: Field Theory as a work in progress', in R. Benson and E. Neveu (eds) *Bourdieu and the Journalistic Field*. Cambridge : Polity, pp. 1–25.

Benson, R. (2006) 'News media as a 'Journalistic Field': what Bourdieu adds to New Institutionalism, and vice versa', *Political Communication*, 23(2), pp. 187–202. Available at: https://doi.org/10.1080/10584600600629802.

Benson, R. (2013) *Shaping Immigration News: A French-American Comparison*. Cambridge: Cambridge University Press.

Benson, R. (2017) 'From heterogeneity to differentiation: searching for a good explanation in a new descriptivist era', in P. Boczkowski and C. W. Anderson (eds) *Remaking the News: Essays on the Future of Journalism Scholarship in the Digital Age*. Cambridge, MA: MIT Press, pp. 27–46.

Benton, T. (2004) 'Critical realism' in M. S. Lewis-Beck, A. Bryman and T. Futing Liao (eds) *The SAGE Encyclopedia of Social Science Research Methods*. Thousand Oaks, CA: SAGE, pp. 222–223. Available at: https://dx.doi.org/10.4135/9781412950589.n197.

Berger, A. A. (2000) *Media and Communication Research Methods: an Introduction to Qualitative and Quantitative Approaches*. Thousand Oaks, CA: SAGE.

Bhaskar, R. (2008) *A Realist Theory of Science*. Abingdon-on-Thames: Routledge.

Bijker, W. E. (1995). *Of bicycles, Bakelites, and Bulbs: Toward a Theory of Sociotechnical Change*. Cambridge, MA: MIT Press.

Bijker, W. E. (2008) 'Technology, social construction of' in W. Donsbach (ed.) *The International Encyclopedia of Communication*. Hoboken, NJ: Wiley. Available at: https://doi.org/10.1002/9781405186407.wbiect025.

Blankespoor E., deHaan E. and Zhu C (2018) 'Capital market effects of media synthesis and dissemination: evidence from robo-journalism', *Review of Accounting Studies*, (23), pp. 1–36. Available at: https://doi.org/10.1007/s11142-017-9422-2.

Blatter, J. K. (2008) 'Case study', in L. M. Given (ed.) *The SAGE Encyclopedia of Qualitative Research Methods*. Thousand Oaks, CA: SAGE, pp. 68–71. Available at: https://dx.doi.org/10.4135/9781412963909.n39.

Bleijenbergh, I. (2010) 'Case selection', in A. J. Mills, G., Durepos and E. Wiebe (eds), *Encyclopedia of Case Study Research*. Thousand Oaks, CA: SAGE, pp. 61–63. Available at: https://dx.doi.org/10.4135/9781412957397.n26.

Blok, A. (2013) 'Actor-network theory', in B. Kaldis (ed.) *Encyclopedia of Philosophy and the Social Sciences*. Thousand Oaks, CA: SAGE, pp. 9–10 Available at: https://dx.doi.org/10.4135/9781452276052.n4.

Blok, A. (2019). 'Actor-network theory, actants' in G. Ritzer (ed.) *The Blackwell Encyclopedia of Sociology*. Hoboken, NJ: Wiley-Blackwell. Available at: https://doi.org/10.1002/9781405165518.wbeosa010.pub2.

Boczkowski, P. J. (2004) 'The mutual shaping of technology and society in Videotex newspapers: beyond the diffusion and social shaping perspectives', *The Information Society*, 20(4), pp. 255–267. Available at: https://doi.org/10.1080/01972240490480947.

Boczkowski, P. J. (2005) *Digitizing the News: Innovation in Online Newspapers*. Cambridge, MA: MIT Press.

Borges-Rey, E. (2016) 'Unravelling data journalism: a study of data journalism practice in British newsrooms', *Journalism Practice*, 10(7), pp. 833–843. Available at: https://doi.org/10.1080/17512786.2016.1159921.

Bourdieu, P. (1992) *Les Règles de l'Art : Genèse et Structure du Champs Littéraire*. Paris: Les Éditions du Seuil.

Bourdieu, P. (1996) *Sur la télévision*. Paris: Les Éditions Raisons d'Agir.

Bourdieu, P. (1997) *Méditations pascaliennes*. Paris: Les Éditions du Seuil.

Bourdieu, P. (2002). *Le Bal des Célibataires : Crise de la Société Paysanne en Béarn*. Paris : Les Éditions du Seuil.

Bourdieu, P. (2005a) 'The Political Field, the Social Science Field, and the Journalistic Field', in R. Benson and E. Neveu (eds) *Bourdieu and the Journalistic Field*. Cambridge: Polity, pp. 29–47.

Bourdieu, P. (2005b) *The Social Structures of the Economy*. Cambridge: Polity.

Bourdieu, P. (2013) 'Séminaires sur le concept de champ, 1972–1975: introduction de Patrick Champagne', *Actes de la Recherche en Sciences Sociales*, 200(5), pp. 4–37.

Boyatzis, R. E. (1998). *Transforming Qualitative Information: Thematic Analysis and Code Development*. Thousand Oaks, CA: SAGE.

Boyle, M. P. and Schmierbach, M. (2019) *Applied Communication Research Methods: Getting Started as a Researcher*. 2nd edition. Abingdon-on-Thames: Routledge.

Boyles, J. L. and Meisinger, J. (2020) 'Automation and adaptation: reshaping journalistic labor in the newsroom library', *Convergence*, 26(1), pp. 178–192. Available at: https://doi.org/10.1177/1354856518772542.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*, 3(2), pp. 77–101. Available at: https://doi.org/10.1191/1478088706qp063oa.

Breed, W. (1955) 'Social control in the newsroom: a functional analysis', *Social Forces*, 33(4), pp. 326–335. Available at: https://doi.org/10.2307/2573002.

Brinkmann, S. and Kvale, S. (2018) *Doing Interviews: Qualitative Research Kit*. Second edition. Thousand Oaks, CA: SAGE.

Broussard, M. (2015) 'Artificial intelligence for investigative reporting', Digital Journalism, 3(6), pp. 814–831. Available at: https://doi.org/10.1080/21670811.2014.985497.

Broussard, M. (2018) *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge, MA: MIT Press.

Buchanan, I. (2010). 'Actor-network theory', in *A Dictionary of Critical Theory*. Available at: https://www.oxfordreference.com/view/10.1093/acref/9780199532919.001.0001/acref-9780199532919-e-9 (Accessed: 8 September 2022).

Callon, M. (1991) 'Techno-economic networks and irreversibility', in: J. Law (ed.) *A Sociology of Monsters: Essays on Power, Technology and Domination*. Abingdon-on-Thames: Routledge, pp. 132–161.

Callon, M. and Latour, B. (1981) 'Unscrewing the big Leviathan or how do actors macrostructure reality and how sociologists help them to do so', in K. Knorr-Cetina and A. Cicourel (eds) *Advances in Social Theory and Methodology: Toward an Integration of Micro and Macro Sociologies*. Abingdon-on-Thames: Routledge, pp. 277–303.

Callon, M. (1984) 'Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay', *The Sociological Review*, 32(1_suppl), pp. 196–233. Available at: https://doi.org/10.1111/j.1467-954X.1984.tb00113.x.

Cardon, D., Cointet, J. and Mazières, A. (2018) 'La revanche des neurones: L'invention des machines inductives et la controverse de l'intelligence artificielle', *Réseaux*, 211, pp. 173–220. Available at : https://doi.org/10.3917/res.211.0173.

Carlson, M. (2018) 'Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism', *New Media & Society*, 20(5), pp. 1755–1772. Available at: https://doi.org/10.1177/1461444817706684.

Carlson, M. (2015) 'The Robotic Reporter', *Digital Journalism*, 3(3), pp. 416–431. Available at: https://doi.org/10.1080/21670811.2014.976412.

Caswell, D. and Dörr, K. (2018) 'Automated Journalism 2.0: Event-driven narratives', *Journalism Practice*, 12(4), pp. 477–496. Available at: https://doi.org/10.1080/17512786.2017.1320773.

Castells, M. (1996) *The Rise of the Networked Society*. Hoboken, NY: Wiley-Blackwell.

Christin, A. (2020) *Metrics at Work: Journalism and the Contested Meaning of Algorithms*. Princeton, NJ: Princeton University Press.

Clark, A. M. (2008) 'Critical realism', in L. M. Given. (ed.) *The SAGE Encyclopedia of Qualitative Research Methods*. Thousand Oaks, CA: SAGE, pp. 168–170. Available at: https://dx.doi.org/10.4135/9781412963909.n87.

Clark, N. (2020) 'What can go wrong when people become interested in the non-human?', in A. Blok, I. Farías and C. Roberts (eds) The Routledge Companion to Actor-Network Theory. Abingdon-on-Thames: Routledge, pp. 158–167.

Clarke, R. (1988) 'Information technology and dataveillance', *Communications of the ACM*, 31(5), pp. 498–512. Available at: https://doi.org/10.1145/42411.42413.

Clerwall, C. (2014) 'Enter the Robot Journalist', *Journalism Practice*, 8(5), pp. 519–531. Available at https://doi.org/10.1080/17512786.2014.883116.

Coddington M. (2015) 'Clarifying journalism's quantitative turn', *Digital Journalism*, 3(3), pp. 331–348. Available at : https://doi.org/10.1080/21670811.2014.976400.

Cohen, S., Li, C., Yang, J. and Yu, C. (2011) 'Computational Journalism: a call to arms to database researchers', *5th Biennial Conference on Innovative Data Systems Research*, Asilomar, CA, 9–12 January. Available at: https://research.google/pubs/pub37381/ (Accessed: 8 September 2022).

Colford, P. (2014) 'A leap forward in quarterly earnings stories', *The Associated Press*, 30 June. Available at: https://blog.ap.org/announcements/a-leap-forward-in-quarterly-earnings-stories (Accessed: July 29 2021).

Collett, M. (2021) 'Hyperdraft: The natural language generation tool powering the ABC Emergency voice app.', *ABC*, 10 June. Available at: https://developers.digital.abc.net.au/2021/06/10/hyperdraft-the-natural-language-generation-tool-powering-the-abc-emergency-voice-app/ (Accessed: 14 August 2022).

Collet, M. (2022) 'Delivering hyperlocal news on demand with ABC Emergency', *ABC*, 11 March,. Available at: https://developers.digital.abc.net.au/2022/03/11/delivering-hyperlocal-news-on-demand-with-abc-emergency/ (Accessed: 14 August 2022).

Cook, K. E. (2008) 'In-depth interview', in L. M. Given (ed.) *The SAGE Encyclopedia of Qualitative Research Methods*. Thousand Oaks, CA: SAGE, p. 423. Available at: https://dx.doi.org/10.4135/9781412963909.n209.

Cools, H., Van Gorp, B. and Opgenhaffen, M. (2022) 'New organizations, different journalistic roles, and innovative projects: how second-generation newsroom innovation labs are changing the news ecosystem', *Journalism Practice*. Available at: https://doi.org/10.1080/17512786.2022.2097940.

Cooren, F. (2009) 'Actor-network theory', in S. W. Littlejohn and K. A. Foss (eds) *Encyclopedia of Communication Theory: Vol. 1*. Thousand Oaks, CA: SAGE, pp. 17–18. Available at: https://dx.doi.org/10.4135/9781412959384.n7.

Cornia, A., Sehl, A. and Nielsen, R. K. (2019) 'Comparing legacy media responses to the changing business of news: Cross-national similarities and differences across media types', *International Communication Gazette*, 81(6–8), pp. 686–706. Available at: https://doi.org/10.1177/1748048518808641.

Cornia, A., Sehl, A. and Nielsen, R. K. (2020) "We no longer live in a time of separation': A comparative analysis of how editorial and commercial integration became a norm', *Journalism*, 21(2), pp. 172–190. Available at: https://doi.org/10.1177/1464884918779919.

Cornia, A., Sehl, A., Levy, D. A. L. and Nielsen, R. K. (2018) *Private sector news, social media distribution, and algorithm change*. Reuters Institute for the Study of Journalism report. Available at: https://reutersinstitute.politics.ox.ac.uk/our-research/private-sector-news-social-media-distribution-and-algorithm-change (Accessed: 8 March 2022).

Corral, D. (2021) 'Algoritmos, precisión y detalle en una cobertura electoral', *RTVE*, 8 April. Available at: https://www.rtve.es/rtve/20210408/inteligencia-artificial-para-cobertura-elecciones/2085044.shtml (Accessed: 12 August 2022).

Couldry, N. (2016) 'Actor-network theory', in K.B. Jensen, E.W. Rothenbuhler, J.D. Pooley and R.T. Craig (eds) *The International Encyclopedia of Communication Theory and Philosophy*. Hoboken, NJ: Wiley. Available at: https://doi.org/10.1002/9781118766804.wbiect012.

Couldry, N. and Mejias, U. A. (2019a) 'Data colonialism: rethinking Big Data's relation to the contemporary subject', *Television & New Media*, 20(4), pp. 336–349. Available at: https://doi.org/10.1177/1527476418796632.

Couldry, N. and Mejias, U. A. (2019b) *The Costs of Connection: How Data is Colonizing Human Life and Appropriating it for Capitalism*. Palo Alto, CA: Stanford University Press.

Couldry, N. (2020) 'Recovering critique in an age of datafication', *New Media & Society*, 22(7), pp. 1135–1151. Available at: https://doi.org/10.1177/1461444820912536.

Crawford, C. S. (2005) 'Actor Network Theory', in G. Ritzer (ed.) *Encyclopedia of Social Theory*. Thousand Oaks, CA: SAGE, pp. 1–3. Available at: https://dx.doi.org/10.4135/9781412952552.n1.

Crossley, N. (2021) 'Habitus/Field' in G. Ritzer (ed.) *The Blackwell Encyclopedia of Sociology*. Hoboken, NJ: Wiley-Blackwell. Available at: https://doi.org/10.1002/9781405165518.wbeosh001.pub2.

Curran, J. (1979) 'Capitalism and Control of the Press, 1800–1975', in J. Curran, M. Gurevitch, and J. Woollacott (eds) *Mass Communication and Society*. Thousand Oaks, CA: SAGE, pp. 195–230.

Czarniawska, B. (2016) 'Actor-network theory' in A. Langley and H. Tsoukas (eds) *The SAGE Handbook of Process Organization Studies*. Thousand Oaks, CA: SAGE, pp. 160–172. Available at: https://dx.doi.org/10.4135/9781473957954.n10.

Danzon-Chambaud, S. (2020) 'Automated fact-checking can catch claims that slip past human checkers. Here are the two ways they work', *Poynter*, 21 August. Available at: https://www.poynter.org/fact-checking/2020/automated-fact-checking-can-catch-claims-that-slip-past-human-checkers-here-are-the-two-ways-they-work/ (Accessed: 10 April 2022).

Danzon-Chambaud, S. (2021a) 'A systematic review of automated journalism scholarship: guidelines and suggestions for future research', *Open Research Europe*, 1(4). Available at: https://doi.org/10.12688/openreseurope.13096.1.

Danzon-Chambaud, S. (2021b) *Covering COVID-19 with Automated News*. Tow Center for Digital Journalism report. Available at: https://www.cjr.org/tow_center_reports/covering-covid-automated-news.php.

Danzon-Chambaud, S. (2021c). *Automated News at the BBC*. BBC Research and Development white paper. Available at: https://www.bbc.co.uk/rd/publications/automated-news-at-bbc-algorithmic-journalism.

Danzon-Chambaud, S. and Cornia, A. (2021) 'Changing or reinforcing the "rules of the game": a Field Theory perspective on the impacts of automated journalism on media practitioners', *Journalism Practice*. Available at: https://doi.org/10.1080/17512786.2021.1919179.

De Maeyer, J., Libert, M., Domingo, D., Heinderyckx, F. and Le Cam, F. (2015) 'Waiting for data journalism: a qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium', *Digital Journalism*, 3(3), pp. 432–446. Available at: https://doi.org/10.1080/21670811.2014.976415.

de-Lima-Santos, M.-F., Schapals, A. K. and Bruns, A. (2021) 'Out-of-the-box versus in-house tools: how are they affecting data journalism in Australia?', *Media International Australia*, 181(1), pp. 152–166. Available at: https://doi.org/10.1177/1329878X20961569.

del Rey García, C. (2020). *Desarrollo y uso de un software de generación de noticias deportivas con sentimiento en la mejora de procesos en el ámbito periodístico*. Bachelor thesis. University Carlos III of Madrid. Available at: http://hdl.handle.net/10016/33042 (Accessed:8 September 2022).

Deuze, M. (2005) 'What is journalism?: professional identity and ideology of journalists reconsidered', *Journalism*, 6(4), pp. 442–464. Available at: https://doi.org/10.1177/1464884905056815.

DeWalt, K. and DeWalt, B. (2011) *Participant observation: a guide for fieldworkers*. 2nd edition. Plymouth: AltaMira Press.

Diakopoulos, N. (2011) 'A Functional Roadmap for Innovation in Computational Journalism', *Nick Diakopoulos: Musings on Media*, 22 April. Available at : https://web.archive.org/web/20220127220537/https://www.nickdiakopoulos.com/2011/04/22/a-functional-roadmap-for-innovation-in-computational-journalism/.

Diakopoulos, N. (2015) 'Algorithmic Accountability', *Digital Journalism*, 3(3), pp. 398–415. Available at: https://doi.org/10.1080/21670811.2014.976411.

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

Diakopoulos, N., Dong, M., Bronner, L. and Bowers, J. (2020) 'Generating location-based news leads for national politics reporting', *Computational + Journalism Symposium 2020*, Boston, MA, 20–21 March. Available at: https://cj2020.northeastern.edu/research-papers/ (Accessed: 14 August 2022).

Díaz-Noci, J. (2020) 'Artificial intelligence systems-aided news and copyright: assessing legal implications for journalism practices', *Future Internet*, 12(5), 85. Available at: https://doi.org/10.3390/fi12050085.

Díaz-Struck, E., Reuter, D., Armendariz, A., Cosic, J., Escudero, J., Fiandor Gutiérrez, M., Torres, M., Kehoe, K., Williams, M., Hassanzade Ajiri, D. and McGoey, S. (2021) 'Pandora Papers: an offshore data tsunami', *International Consortium of Investigative Journalists*, October 3. Available at: https://www.icij.org/investigations/pandora-papers/about-pandora-papers-leak-dataset/ (Accessed: 8 September 2022).

Díaz-Struck, E., Romera, P. and Ledésert, S. (2020) 'How we mined more than 715,000 Luanda Leaks records', *International Consortium of Investigative Journalists*, 3 February. Available at: https://www.icij.org/investigations/luanda-leaks/how-we-mined-more-than-715000-luanda-leaks-records/ (Accessed: July 29 2021).

Dierickx, L. (2020) 'Journalists as end-users: quality management principles applied to the design process of news automation', *First Monday*, 25(4). Available at: https://doi.org/10.5210/fm.v25i4.10558

Dierickx, L. (2021) 'News automation, materialities, and the remix of an editorial process', *Journalism*. Available at: https://doi.org/10.1177/14648849211023872.

Döllerer, L. (2021) 'We used OpenAI's GPT-3 to keep teenagers informed', *Lukas Döllerer: Personal Webpage*, 13 April. Available at: https://lukas-doellerer.de/en/post/openai-gpt3-news-summarizer/ (Accessed: 20 August 2022).

Dörr, K. N. (2016) 'Mapping the field of Algorithmic Journalism', *Digital Journalism*, 4(6), pp. 700–722. Available at: https://doi.org/10.1080/21670811.2015.1096748.

Dörr, K. N. and Hollnbuchner, K. (2017) 'Ethical challenges of algorithmic journalism', *Digital Journalism*, 5(4), pp. 404–419. Available at: https://doi.org/10.1080/21670811.2016.1167612.

Elger, T. (2010) 'Critical realism' in A. J. Mills, G., Durepos and E. Wiebe (eds). *Encyclopedia of Case Study Research*. Thousand Oaks, CA: SAGE, pp. 254–257. Available at: https://dx.doi.org/10.4135/9781412957397.n95.

Emerging Technology from the arXiv (2017) 'How Reuters's revolutionary AI system gathers global news', *MIT Technology Review*, 27 November. Available at: https://www.technologyreview.com/2017/11/27/241656/how-reuterss-revolutionary-ai-system-gathers-global-news/ (Accessed: 14 August 2022).

English, P. (2016) 'Mapping the Sports Journalism Field: Bourdieu and broadsheet newsrooms', *Journalism*, 17(8): pp. 1001–1017. Available at: https://doi.org/10.1177/1464884915576728.

Epstein, E. J. (1973) *News from Nowhere: Television and the News*. New York, NY: Random House.

Ezzy, D. (2002) *Qualitative Analysis*. Abingdon-on-Thames: Routledge.

Falk Eriksen, M. (2018) *When Robots Become Colleagues: A Case Study of the Danish News Agency Ritzau*. Unpublished Master thesis. Aarhus University and Hamburg University.

Fanta, A. (2017) *Putting Europe's Robots on the Map: Automated Journalism in News Agencies*. Reuters Institute for the Study of Journalism fellowship paper. Available at: https://reutersinstitute.politics.ox.ac.uk/our-research/putting-europes-robots-map-automated-journalism-news-agencies (Accessed: July 29 2021).

Fell, J. (2021a) 'Aggregating emergency alerts', *ABC*, 9 April. Available at: https://developers.digital.abc.net.au/2021/04/09/aggregating-emergency-alerts/ (Accessed: 16 August 2022).

Fell, J. (2021b) 'A framework for conversational apps.', *ABC*, 3 November. Available at: https://developers.digital.abc.net.au/2021/11/03/a-framework-for-conversational-apps/ (Accessed: 16 August 2022).

Ferrer-Conill, R. (2017) 'Quantifying journalism? A study on the use of data and gamification to motivate journalists', *Television & New Media*, 18(8), pp. 706–720. Available at: https://doi.org/10.1177/1527476417697271.

Ferrucci, P. and Eldridge II, S. A. (eds) (2022) *The Institutions Changing Journalism: Barbarians Inside the Gate*. Edited volume. Abingdon-on-Thames: Routledge.

Fishman, M. (1980) *Manufacturing the News*. Austin, TX: University of Texas Press.

Fletcher, R. and Jenkins, J. (2019) *Polarisation and the News Media in Europe*. European Parliament's Panel for the Future of Science and Technology report. Available at https://reutersinstitute.politics.ox.ac.uk/our-research/polarisation-and-news-media-europe (Accessed: 8 September 2022).

Flew, T., Spurgeon, C., Daniel, A. and Swift, A. (2012) 'The promise of computational journalism', *Journalism Practice*, 6(2), pp. 157–171. Available at: https://doi.org/10.1080/17512786.2011.616655.

Ford, H. and Hutchinson, J. (2019) 'Newsbots that mediate journalist and audience relationships', *Digital Journalism*, 7(8): pp. 1013–1031. Available at: https://doi.org/10.1080/21670811.2019.1626752.

Galindo, G. (2020) 'XR Belgium posts deepfake of Belgian premier linking Covid-19 with climate crisis', *The Brussels Times*, 14 April. Available at: https://brusselstimes.com/health/106320/xr-belgium-posts-deepfake-of-belgian-premier-linking-covid-19-with-climate-crisis (Accessed: 15 December 2021).

Gans, H. J. (1979) *Deciding What's News: A Study of CBS Evening News, NBC Nightly News, Newsweek, and Time*. 2004 Edition. Evanston, IL: Northwestern University Press.

García-Perdomo, V., Montaña-Niño, S. X. and Magana, M. I. (2022) *Envisioning automated journalism in Latin American media* [Presentation]. ICA preconference Understanding the Dynamics of (Ir)Responsible AI in Journalism and Algorithmically Shaped News Flows, Paris. 26 May.

Gebru, T. (2019) 'Race and gender', *arXiv*. Available at: https://arxiv.org/abs/1908.06165 (Accessed: October 16 2021).

Gee, S. and Prior, F. (2018) 'Backstory: ABC News bot partners with Hearken platform in world-first project to tap SA voters' curiosity', *ABC*, 29 March. Available at: https://www.abc.net.au/news/redirects/backstory/digital/2018-03-29/abc-news-bot-and-hearken-curious-platform-during-sa-election/9591966 (Accessed: 14 August 2022).

Gerring, J. (2007). *Case Study Research: Principles and Practices*. Cambridge, UK: Cambridge University Press.

Giddens, A. (1984) *The Constitution of Society: Outline of the Theory of Structuration*. Cambridge: Polity

Gillespie, T. (2014) 'The relevance of algorithms', in T. Gillespie, P. J. Boczkowski and K. A. Foot (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: MIT Press, pp. 167–193.

Gordon, R. (2019) 'Better fact-checking for fake news', *MIT News*, 17 October. Available at: https://news.mit.edu/2019/better-fact-checking-fake-news-1017 (Accessed: July 29 2021).

Graefe, A. (2016) *Guide to Automated Journalism*. Tow Center for Digital Journalism report. Available at: https://academiccommons.columbia.edu/doi/10.7916/D80G3XDJ (Accessed: 29 July 2021).

Graefe, A., Haim, M., Haarmann, B. and Brosius, H. B. (2018) 'Readers' perception of computer-generated news: credibility, expertise, and readability', *Journalism*, 19(5), pp. 595–610. Available at: https://doi.org/10.1177/1464884916641269.

Graefe, A. and Bohlken, N. (2020) 'Automated journalism: a meta-analysis of readers' perceptions of human-written in comparison to automated news', *Media and Communication*, 8(3), pp. 50–59. Available at: https://doi.org/10.17645/mac.v8i3.3019.

Graves, L. (2018) *Understanding the Promise and Limits of Automated Fact-Checking*. Reuters Institute for the Study of Journalism factsheet. Available at: https://reutersinstitute.politics.ox.ac.uk/our-research/understanding-promise-and-limits-automated-fact-checking (Accessed: 29 July 2021).

Green, T. (2019) 'Salco part 2: automating our local elections coverage', *BBC News Labs*, August 7. Available at: https://bbcnewslabs.co.uk/news/2019/salco-local-elections/ (Accessed: 12 August 2022).

Guardian reporters and Guardian readers (2009) 'MPs' expenses—what you've discovered', *The Guardian*, 16 December. Available at: https://www.theguardian.com/politics/2009/dec/16/mps-expenses-what-we-learned (Accessed: 27 June 2022).

Gurrin, C., Smeaton, A. F. and Doherty, A. R. (2014) 'LifeLogging: personal big data', *Foundations and Trends in Information Retrieval*, 8(1), pp. 1–125. http://dx.doi.org/10.1561/1500000033.

Gynnild, A. (2014) 'Journalism innovation leads to innovation journalism: the impact of computational exploration on changing mindsets', *Journalism*, 15(6), pp. 713–730. Available at: https://doi.org/10.1177/1464884913486393.

Haapanen, L. (2020) *Media councils and self-regulation in the emerging era of news automation*. Council for Mass Media in Finland report. Available at: http://urn.fi/URN:NBN:fi:jyu-202012297409 (Accessed: 18 March 2022).

Hagar, N., Diakopoulos, N. and DeWilde, B. (2022) 'Anticipating attention: on the predictability of news headline tests', *Digital Journalism*, 10(4), pp. 647–668. Available at: https://doi.org/10.1080/21670811.2021.1984266.

Haim, M. and Graefe, A. (2017) 'Automated news: better that expected?' *Digital Journalism*, 5(8), pp. 1044–1059. Available at: https://doi.org/10.1080/21670811.2017.1345643.

Hallin, D. C. and Mancini, P. (2004) *Comparing Media Systems: Three Models of Media and Politics*. Cambridge: Cambridge University Press.

Hallin, D. C. and Mancini, P. (eds) (2011) *Comparing Media Systems Beyond the Western World*. Cambridge: Cambridge University Press.

Hamilton, J. T. and Turner, F. (2009) *Accountability through Algorithm: Developing the Field of Computational Journalism*. Stanford's Center For Advanced Study in the Behavioral Sciences Summer Workshop report. Available at: https://fredturner2022.sites.stanford.edu/sites/g/files/sbiybj27111/files/media/file/hamilton-turner-accountability-through-algorithm-white-paper.pdf (Accessed: 9 September 2022).

Hamilton, J. T. (2018) *Democracy's Detectives: the Economics of Investigative Journalism*. Cambridge, MA: Harvard University Press.

Hansen, A. and Machin, D. (2013) *Media and Communication Research Methods*. London: Palgrave Macmillan.

Harwell, D. and Miroff, N. (2018) 'ICE just abandoned its dream of 'extreme vetting' software that could predict whether a foreign visitor would become a terrorist', *The Washington Post*, 17 May. Available at: https://www.washingtonpost.com/news/the-switch/wp/2018/05/17/ice-just-abandoned-its-dream-of-extreme-vetting-software-that-could-predict-whether-a-foreign-visitor-would-become-a-terrorist/ (Accessed: 25 August 2022).

Hassard, J. and Alcadipani, R. (2010) 'Actor-network theory', in A. J. Mills, G. Durepos and E. Wiebe (eds) *Encyclopedia of Case Study Research*. Thousand Oaks, CA: SAGE, pp. 9–12. Available at: https://dx.doi.org/10.4135/9781412957397.n4.

Haugeland, J. (1985) *Artificial Intelligence: the Very Idea*. Cambridge, MA: MIT Press.

Heiselberg, L., Blom, J. N. and van Dalen, A. (2022) 'Automated news reading in the neural age: audience reception and perceived credibility of a news broadcast read by a neural voice', *Journalism Studies*, 23(8), pp. 896–914. Available at: https://doi.org/10.1080/1461670X.2022.2052346.

Helberger, N. and Diakopoulos, N. (2022) 'The European AI Act and how it matters for research into AI in media and journalism', *Digital Journalism*. Available at: https://doi.org/10.1080/21670811.2022.2082505.

Hemmingway, E. (2005) 'PDP, the news production network and the transformation of news', *Convergence*, 11(3), pp. 8–27. Available at: https://doi.org/10.1177/1354856505011100302.

Hennis, W. (1998) 'The media as a cultural problem: Max Weber's sociology of the press', *History of the Human Sciences*, 11(2), pp. 107–110. Available at https://doi.org/10.1177/095269519801100206.

Heravi, B. R. and Harrower, N. (2016) 'Twitter journalism in Ireland: sourcing and trust in the age of social media', *Information, Communication & Society*, 19(9), pp. 1194-1213. Available at: https://doi.org/10.1080/1369118X.2016.1187649.

Heravi, B. R. (2019) '3Ws of data journalism education', *Journalism Practice*, 13(3), pp. 349–366. Available at: https://doi.org/10.1080/17512786.2018.1463167.

Hermida, A. and Thurman, N. (2008) 'A clash of cultures: the integration of user-generated content within professional journalistic frameworks at British newspaper websites', *Journalism Practice*, 2(3), pp. 343–356. Available at: https://doi.org/10.1080/17512780802054538.

Houston, S. (2014) 'Critical realism', in D. Coghlan and M. Brydon-Miller (eds) *The SAGE Encyclopedia of Action Research: Vol. 2*. Thousand Oaks, CA: SAGE, pp. 220–222. Available at: https://dx.doi.org/10.4135/9781446294406.n95.

Hovland, C. I. and Weiss, W. (1951) 'The influence of source credibility on communication effectiveness', *Public Opinion Quarterly*, 15(4), pp. 635–650. Available at: https://doi.org/10.1086/266350.

Howard, P. (2003) 'McLuhan, Marshall', in S. Jones (ed.) *Encyclopedia of New Media*. Thousand Oaks, CA: SAGE, pp. 305–307. Available at: https://doi.org/10.4135/9781412950657.n157.

Hughes, T. P. (1994) 'Technological momentum', in M. R. Smith and L. Marx (eds.) *Does Technology Drive History?* Cambridge, MA: MIT Press, pp. 99–113.

Humphreys, L. (2010) 'Technological determinism', in S. H. Priest *Encyclopedia of Science and Technology Communication*. Thousand Oaks, CA: SAGE, pp. 869–872. Available at: https://doi.org/10.4135/9781412959216.n293.

Hussey, P., Das, S., Farrell, S., Ledger, L. and Spencer, A. (2021) 'A knowledge graph to understand nursing Big Data: case example for guidance', *Journal of Nursing Scholarship*, 53(3), pp. 323–332. Available at: https://doi.org/10.1111/jnu.12650.

Hutton, R. (2019) 'Stories by numbers: experimenting with semi-automated journalism', *BBC News Labs*, 22 March. Available at: https://bbcnewslabs.co.uk/news/2019/stories-by-numbers/ (Accessed: 12 August 2022).

Innis, H. (1950) *Empire and Communications*. Oxford: Oxford University Press.

James, L. M. and Vo, H. T. (2010) 'Hawthorne effect', in N. J. Salkind (ed.) *Encyclopedia of Research Design: Vol. 1*. Thousand Oaks, CA: SAGE, pp. 561–563. Available at: https://dx.doi.org/10.4135/9781412961288.n174.

Jenkins, R. (2005) 'Bourdieu, Pierre', in G. Ritzer (ed.) *Encyclopedia of Social Theory: Vol. 1* Thousand Oaks, CA: SAGE, pp. 67–71. Available at: https://doi.org/10.4135/9781412952552.n31.

Jia, C. and Johnson, T. (2021) 'Source credibility matters: does automated journalism inspire selective exposure?' *International Journal Of Communication*, 15(2021), pp. 3760–3781. Available at: https://ijoc.org/index.php/ijoc/article/view/16546/3529.

Jones, B. and Jones, R. (2019) 'Public service chatbots: automating conversation with BBC News', *Digital Journalism*, 7(8), pp. 1032–1053. Available at: https://doi.org/10.1080/21670811.2019.1609371.

Jones, R. and Jones, B. (2019) 'Atomising the news: the (in)flexibility of structured journalism', *Digital Journalism*, 7(8), pp. 1157–1179. Available at: https://doi.org/10.1080/21670811.2019.1609372.

Jung, J., Song, H., Kim, Y., Im, H. and Oh, S. (2017) 'Intrusion of software robots into journalism: the public's and journalists' perceptions of news written by algorithms and human journalists', *Computers in Human Behavior*, 71(2017), pp. 291–298. Available at: https://doi.org/10.1016/j.chb.2017.02.022.

Karlsen, J. and Stavelin, E. (2014) 'Computational journalism in Norwegian newsrooms', *Journalism Practice*, 8(1), pp. 34–48. Available at: https://doi.org/10.1080/17512786.2013.813190.

Kanerva, J., Rönnqvist, S., Kekki, R., Salakoski, T. and Ginter, F. (2019) 'Template-free data-to-text generation of Finnish sports news', *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, Turku, 30 September–2 October. Linköping: Linköping University Electronic Press, pp. 242–252. Available at: https://ep.liu.se/ecp/167/025/ecp19167025.pdf (Accessed: 9 September 2022).

Karlsefni, T. (2019) *Nicholas Cage, Sound of Music (deepfake)*. 10 October. Available at: https://www.youtube.com/watch?v=MHkZEpfUnAA (Accessed: 15/12/2021).

Kim, D. and Kim, S. (2017) 'Newspaper companies' determinants in adopting robot journalism', *Technological Forecasting and Social Change*, 117(2017), pp. 184–195. Available at: https://doi.org/10.1016/j.techfore.2016.12.002.

Kim, D. and Kim, S. (2018) 'Newspaper journalists' attitudes towards robot journalism', *Telematics and Informatics*, 35(2), pp. 340–357. Available at: https://doi.org/10.1016/j.tele.2017.12.009.

Kim D. and Lee, J. (2019) 'Designing an algorithm-driven text generation system for personalized and interactive news reading', *International Journal of Human–*

Computer Interaction, 35(2), pp. 109–122. Available at: https://doi.org/10.1080/10447318.2018.1437864.

Kofman, A. (2022) 'On Bruno Latour (1947–2022): the world was his laboratory', *n+1*, 13 October. Available at: https://www.nplusonemag.com/online-only/online-only/on-bruno-latour-1947-2022/ (Accessed: 20 December 2022).

Kozhevnikov, A. and Vincent, S. (2019) 'Critical realism', in P. Atkinson (ed.) *SAGE Research Methods Foundations*. Thousand Oaks, CA: SAGE. Available at: https://dx.doi.org/10.4135/9781526421036782465.

Kozinets, R. V. (2016) 'Netnography', in G. Ritzer (ed.) *The Blackwell Encyclopedia of Sociology*. Hoboken, NJ: Wiley-Blackwell. Available at: https://doi.org/10.1002/9781405165518.wbeos0782.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012) 'ImageNet classification with deep convolutional neural networks', *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Granada, 12–17 December 2011. Available at: https://papers.nips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html (Accessed: 9 September 2022).

Kunert, J. (2020) 'Automation in sports reporting: strategies of data providers, software providers, and media outlets', *Media and Communication*, 8(3), pp. 5–15. Available at: https://doi.org/10.17645/mac.v8i3.2996.

L.A. Times: L.A. Now (2017) 'Please note: We just deleted an automated tweet saying there was a 6.8 earthquake in Isla Vista. That earthquake happened in 1925.' [Twitter] June 22. Available at: https://twitter.com/LANow/status/877677781089304576 (Accessed: 27 June 2022).

Lapadat, J. C. (2010) 'Thematic analysis', in A. J. Mills, G., Durepos and E. Wiebe (eds) *Encyclopedia of Case Study Research*. Thousand Oaks: SAGE, pp. 926–927. Available at: https://dx.doi.org/10.4135/9781412957397.n342.

Latour, B. (1992) 'Where are the missing masses? The sociology of a few mundane artifacts', in W. E. Bijker and J. Law (eds) *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge, MA: The MIT Press, pp. 225–258.

Latour, B. (1988) *The Pasteurization of France*. Cambridge, MA: Harvard University Press.

Latour, B. (2005) *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford: Oxford University Press.

Law, J. (1984) 'On the methods of long-distance control: vessels, navigation and the Portuguese route to India', *The Sociological Review*, 32(1_suppl), pp. 234–263. Available at: https://doi.org/10.1111/j.1467-954X.1984.tb00114.x.

Law, J. (1992) 'Notes on the theory of the actor-network: ordering, strategy, and heterogeneity', Systems Practice, 5(1992), pp. 379–393. Available at: https://doi.org/10.1007/BF01059830.

Lemelshtrich Latar, N. (2015) 'The robot journalist in the age of social physics: the end of human journalism?', in G. Einav (ed.) *The New World of Transitioned Media*. Cham: Springer, pp. 65–80. Available at: https://doi.org/10.1007/978-3-319-09009-2_6.

Lemelshtrich Latar, N. (2018) 'Robot journalism' in N. Lemelshtrich Latar (ed.) *Robot Journalism: Can Human Journalism Survive?* Singapore: World Scientific, pp. 29–40.

Leonard, M. (2003) 'Interviews', in R. L. Miller and J. Brewer (eds) *The A–Z of Social Research*. Thousand Oaks: SAGE, pp. 167–171. Available at: https://doi.org/10.4135/9780857020024.n59.

Leppänen, L. (2019) *Automated news text production: where we are and where we might be heading* [Presentation]. News Automation at Work PhD workshop, Dublin, 4–5 November.

Leppänen, L., Tuulonen, H. and Sirén-Heikel, S. (2020) 'Automated journalism as a source of and a diagnostic device for bias in reporting', *Media and Communication*, 8 (3), pp. 39–49. Available at: https://doi.org/10.17645/mac.v8i3.3022.

Levy, S. (2012) 'Can an algorithm write a better news story than a human reporter?' *Wired*, 24 April. Available at: https://www.wired.com/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter/ (Accessed: 9 September 2022).

Lewis, S. C. and Usher, N. (2013) 'Open source and journalism: toward new frameworks for imagining news innovation', *Media, Culture & Society*, 35(5), pp. 602–619. Available at: https://doi.org/10.1177/0163443713485494.

Lewis, S. C. and Westlund, O. (2015a) 'Big Data and journalism', *Digital Journalism*, 3(3), pp. 447–466. Available at: https://doi.org/10.1080/21670811.2014.976418.

Lewis, S. C. and Westlund, O. (2015b) 'Actors, actants, audiences, and activities in cross-media news work', *Digital Journalism*, 3(1), pp. 19–37. Available at: https://doi.org/10.1080/21670811.2014.927986.

Lewis, S. C. and Usher, N. (2016) 'Trading zones, boundary objects, and the pursuit of news innovation: A case study of journalists and programmers', *Convergence*, 22(5), pp. 543–560. Available at: https://doi.org/10.1177/1354856515623865.

Lewis, S. C., Sanders, A. K. and Carmody, C. (2019) 'Libel by algorithm? Automated journalism and the threat of legal liability', *Journalism & Mass Communication Quarterly*, 96(1), pp. 60–81. Available at: https://doi.org/10.1177/1077699018755983.

Lindblom, T., Lindell, J. and Gidlund, K. (2022) 'Digitalizing the Journalistic Field: journalists' views on changes in journalistic autonomy, capital and habitus', *Digital Journalism*. Available at: https://doi.org/10.1080/21670811.2022.2062406.

Lindén, C.-G. (2017a) 'Decades of automation in the newsroom: why are there still so many jobs in journalism?' *Digital Journalism*, 5(2), pp. 123–140. Available at: https://10.1080/21670811.2016.1160791.

Lindén, C.-G. (2017b) 'Algorithms for journalism: the future of news work', *The Journal of Media Innovations*, 4(1) , pp. 60–76. Available at: https://doi.org/10.5617/jmi.v4i1.2420.

Lindén, C.-G. and Dierickx, L. (2019) 'Robot journalism: the damage done by a metaphor', *Unmediated: Journal of Politics and Communication*, 2, pp. 152–155. Available at: https://researchportal.helsinki.fi/sv/publications/robot-journalism-the-damage-done-by-a-metaphor-2 (Accessed: 9 September 2022).

Lindén, C.-G. and Tuulonen, H. (eds.) (2019) *News Automation: The Rewards, Risks and Realities of 'Machine Journalism'*. WAN-IFRA report. Available at: http://hdl.handle.net/10138/305568 (Accessed: 9 September 2022).

Lindén, C.-G., Tuulonen, H.; Niemi, L., Grönlund, M., Lehtisaari, K. and Villi, M. (2022) *Road to freedom: news media breaking free from Silicon Valley's grip*. University of Helsinki's Swedish School of Social Science report. Available at: http://hdl.handle.net/10138/340138 (Accessed: 20 May 2022).

Liu, B. and Wei, L. (2019) 'Machine authorship *in situ*', *Digital Journalism*, 7(5), pp. 635–657. Available at: https://doi.org/10.1080/21670811.2018.1510740.

Liu, X., Nourbakhsh, A., Li, Q., Shah, S., Martin, R. and Duprey, J. (2017) 'Reuters Tracer: toward automated news production using large scale social media data', *2017 IEEE International Conference on Big Data* (BIGDATA), Boston, MA, 11–14 December. Available at: https://doi.org/10.1109/BigData.2017.8258082.

Lokot, T. and Diakopoulos, N. (2016) 'News bots: automating news and information dissemination on Twitter', *Digital Journalism*, 4(6), pp. 682–699. Available at: https://www.doi.org/10.1080/21670811.2015.1081822.

Lupton, D. (2016) *The Quantified Self: A Sociology of Self-Tracking*. Cambridge: Polity.

Macgregor, M. (2021) 'Responsible AI at the BBC: our machine learning engine principles', *BBC*, May. Available at: https://www.bbc.co.uk/rd/publications/responsible-ai-at-the-bbc-our-machine-learning-engine-principles (Accessed: 9 September 2022).

Manning White, D. (1950) 'The "gate keeper": a case study in the selection of news', *Journalism & Mass Communication Quarterly*, 27(4), pp. 383–390. Available at: https://doi.org/10.1177/107769905002700403.

Menke, M., Kinnebrock, S., Kretzschmar, S., Aichberger, I., Broersma, M., Hummel, R., Kirchhoff, S., Prandner, D., Ribeiro, N. and Salaverría, R. (2018) 'Convergence culture in European newsrooms', *Journalism Studies*, 19(6), pp. 881–904. Available at: https://doi.org/10.1080/1461670X.2016.1232175.

Marchand, V. (2019) 'Tamedia utilisera le robot "Tobi" pour couvrir les élections fédérales', *Cominmag*, 7 October. Available at: https://cominmag.ch/tamedia-utilisera-le-robot-tobi-pour-couvrir-les-elections-federales/ (Accessed: 14 August 2022).

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

Marconi, F. and Siegman, A. (2017). *The Future of Augmented Journalism*. Associated Press report. Available at: https://www.ap.org/assets/files/2017_ai_guide.pdf (Accessed: 20 August 2022).

Mari, W. (2018) 'Technology in the newsroom: adoption of the telephone and the radio car from c. 1920 to 1960', *Journalism Studies*, 19 (9), pp. 1366–1389. Available at: https://doi.org/10.1080/1461670X.2016.1272432.

Mari, W. (2019) *A Short History of Disruptive Journalism Technologies: 1960–1990*. Abingdon-on-Thames: Routledge

Mari, W. (2021) 'Staying cool: the impact of air conditioning on news work and the modern newsroom', *Journalism Practice*. Available at: https://doi.org/10.1080/17512786.2021.1919544.

Mason, J. (2002) *Qualitative Researching*. 2nd edition. Thousand Oaks, CA: SAGE.

Mateus, S. (2020) 'Medium', in D. Merskin (ed.) *The SAGE International Encyclopedia of Mass Media and Society: Vol. 5*. Thousand Oaks, CA: SAGE, pp. 1089–1091. Available at: https://doi.org/10.4135/9781483375519.

Mayer-Schönberger, V. and Cukier, K. (2014) *Big Data: A Revolution that will Transform How we Live, Work, and Think*. Boston, MA: Houghton Mifflin Harcourt.

McLuhan, M. (1962) *The Gutenberg Galaxy: The Making of Typographic Man*. Toronto, ON: University of Toronto Press.

McLuhan, M. (1964) *Understanding Media: The Extensions of Man*. New York, NY: McGraw Hill.

Meij, E. (2019) 'Understanding news using the Bloomberg Knowledge Graph' [Speakerdeck], 14 May. Available at: https://speakerdeck.com/emeij/understanding-news-using-the-bloomberg-knowledge-graph (Accessed: 12 August 2022).

Melin, M., Bäck. A., Södergård, C., Munezero, M. D., Leppänen, L. and Toivonen, H. (2018) 'No landslide for the human journalist: an empirical study of computer-generated election news in Finland', *IEEE Access*, 6(2018), pp. 43356–43367. Available at: https://doi.org/10.1109/ACCESS.2018.2861987.

Metz, C. (2018) 'Why A.I. and cryptocurrency are making one type of computer chip scarce', *The New York Times*, 8 May. Available at:

https://www.nytimes.com/2018/05/08/technology/gpu-chip-shortage.html
(Accessed: 16 October 2021).

Meyen, M. and Riesmeyer, C. (2012) 'Service providers, sentinels, and traders: journalists'
role perceptions in the early twenty-first century', *Journalism Studies*, 13(3), pp. 386–
401. Available at: https://doi.org/10.1080/1461670X.2011.602909.

Meyer, J. W. and Rowan, B. (1977) 'Institutionalized organizations: formal structure as
myth and ceremony', *American Journal of Sociology*, 83(2), pp. 340–363. Available at:
http://www.jstor.org/stable/2778293 (Accessed: 9 September 2022).

Meyer, P. (1973) *Precision Journalism: A Reporter's Introduction to Social Science Methods*.
Bloomington: Indiana University Press.

Meyer, P. (1988) 'Defining and measuring credibility of newspapers: developing an index',
Journalism & Mass Communication Quarterly, 65(3), pp. 567–574. Available at:
https://doi.org/10.1177/107769908806500301.

Michael, M. (2017a) *Actor-Network Theory: Trials, Trails and Translations*. Thousand Oaks,
CA: SAGE.

Michael, M. (2017b) 'Actor Network Theory', in B.S. Turner (ed.) *The Wiley-Blackwell
Encyclopedia of Social Theory*. Hoboken, NJ: Wiley-Blackwell. Available at:
https://doi.org/10.1002/9781118430873.est0002.

Mills, J. and Wagemans, A. (2021) 'Media labs: Constructing journalism laboratories,
innovating the future: how journalism is catalysing its future processes, products
and people', *Convergence*, 27(5), pp. 1462–1487. Available at:
https://doi.org/10.1177/1354856521994453.

Milosavljević, M. and Vobič, I. (2021) "Our task is to demystify fears': analysing newsroom
management of automation in journalism', *Journalism*, 22(9), pp. 2203–2221.
Available at: https://doi.org/10.1177/1464884919861598.

Molotch, H. and Lester, M. (1974) 'News as purposive behavior: on the strategic use of
routine events, accidents and scandals', *American Sociological Review*, 39(1), pp. 101–
113. Available at: https://doi.org/10.2307/2094279.

Molumby, C. (2019) 'Trees, machines and local journalism', *BBC News Labs*, 28 November.
Available at: https://bbcnewslabs.co.uk/news/2019/salco-trees/ (Accessed: 12 August
2022).

Molumby, C. and Whitwell, J. (2019) 'General election 2019: semi-automation makes it a
night of 689 stories', *BBC News Labs*, 13 December. Available at:
https://bbcnewslabs.co.uk/news/2019/salco-ge/ (Accessed: 12 August 2022).

Montal, T. and Reich, Z. (2017) 'I, robot. You, journalist. Who is the author?' *Digital
Journalism*, 5(7), pp. 829–849. Available at:
https://doi.org/10.1080/21670811.2016.1209083.

Moravec, V., Macková, V., Sido, J. and Ekštein, K. (2020) 'The robotic reporter in the Czech news agency: automated journalism and augmentation in the newsroom', *Communication Today*, 11(1), pp. 36–52. Available at: https://www.communicationtoday.sk/the-robotic-reporter-in-the-czech-news-agency-automated-journalism-and-augmentation-in-the-newsroom/ (Accessed: 20 May 2022).

Morey Hawkins, J. (2017) 'Thematic analysis', in M. Allen (ed.) *The SAGE Encyclopedia of Communication Research Methods: Vol. 4*. Thousand Oaks, CA: SAGE, pp. 1757–1760. Available at: https://dx.doi.org/10.4135/9781483381411.n624.

Muir, L. (2020) 'Information society', in D. Merskin (ed.) *The SAGE International Encyclopedia of Mass Media and Society: Vol. 5*. Thousand Oaks, CA: SAGE, pp. 798–800. Available at: https://doi.org/10.4135/9781483375519.

Mullin, B. (2015) 'With new product, Automated Insights hopes to make 'robot journalism' cheaper and more plentiful', *Poynter*, 20 October. Available at: https://www.poynter.org/newsletters/2015/with-new-product-automated-insights-hopes-to-make-robot-journalism-cheaper-and-more-plentiful/ (Accessed: 11 August 2022).

Murdock, G. and Golding, P. (1977) 'Capitalism, communications and class relations', in J. Curran, M. Gurevitch, and J. Woollacott (eds) *Mass Communication and Society*. Thousand Oaks, CA: SAGE, pp. 12–43.

Napoli, P. M. (2014) 'Automated media: an Institutional Theory perspective on algorithmic media production and consumption', *Communication Theory*, 24(3), pp. 340–360. Available at: https://doi.org/10.1111/comt.12039.

Narrativa, (no date 1) *Narrativa's Advanced Natural Language Technology*. Available at: https://www.narrativa.com/narrativas-technology-stack-nlg-nlp-and-data-analysis-and-augmentation/ (Accessed July 29 2021).

Narrativa Staff (no date 2) *ANSA Agency uses Narrativa AI for news on COVID-19*. Available at: https://www.narrativa.com/ansa-agency-uses-narrativa-ai-for-news-on-covid-19/ (Accessed: 12 August 2022).

Nechushtai, E. (2018) 'From Liberal to Polarized Liberal? Contemporary U.S. news in Hallin and Mancini's typology of news systems', *The International Journal of Press/Politics*, 23(2), pp. 183–201. Available at: https://doi.org/10.1177/1940161218771902.

Nelson, J. L. (2021) *Imagined Audiences: How Journalists Perceive and Pursue the Public*. Oxford: Oxford University Press.

Neveu, E. (2005) 'Bourdieu, the Frankfurt School, and cultural studies: on some misunderstandings' in R. Benson and E. Neveu (eds) *Bourdieu and the Journalistic Field*. Cambridge: Polity, pp. 1–25.

Neveu, E. (2007) 'Pierre Bourdieu : sociologist of media, or sociologist for media scholars?' *Journalism Studies*, 8(2), pp. 335–347. Available at: https://doi.org/10.1080/14616700601149026.

Nielsen, R. K. (2012) 'How newspapers began to blog', *Information, Communication & Society*, 15(6), pp. 959–978. Available at: https://doi.org/10.1080/1369118X.2012.694898.

Nielsen, R. K. (2016) 'The business of news', in T. Witschge, C. W. Anderson and D. Domingo (eds) *The SAGE Handbook of Digital Journalism*. Thousand Oaks, CA: SAGE, pp. 51–67. Available at: https://doi.org/10.4135/9781473957909.n4.

Nielsen, R. K. and Ganter, S. A. (2022) *The Power of Platforms: Shaping Media and Society*. Oxford: Oxford University Press.

Nikolova, N. (2008) 'Actor-network theory', in S. R. Clegg and J. R. Bailey (eds) *International Encyclopedia of Organization Studies*. Thousand Oaks, CA: SAGE, pp. 28–31 Available at: https://dx.doi.org/10.4135/9781412956246.n9.

O'Neill, D. and O'Connor, C. (2008) 'The passive journalist: how sources dominate local news', *Journalism Practice*, 2(3), pp. 487–500. Available at: https://doi.org/10.1080/17512780802281248.

O'Reilly, K. (2011) *Ethnographic Methods*. 2nd edition. Abingdon-on-Thames: Routledge.

O'Reilly, K. (2009) 'Inductive and deductive', in K. O'Reilly (ed.) *Key Concepts in Ethnography*. Thousand Oaks: SAGE, pp. 104–109. Available at: https://dx.doi.org/10.4135/9781446268308.n19.

Olteanu, A., Garcia-Gathright, J., de Rijke, M. and Ekstrand, Mi. D. (2019) 'FACTS-IR: fairness, accountability, confidentiality, transparency, and safety in information retrieval', *ACM SIGIR Forum*, 53(2), pp. 20–43. Available at: https://doi.org/10.1145/3458553.3458556.

Örnebring, H. (2010) 'Technology and journalism-as-labour: historical perspectives', *Journalism*, 11(1), pp. 57–74. Available at: https://doi.org/10.1177/1464884909350644.

Parasie, S. and Dagiral, E. (2013) 'Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in Chicago', *New Media & Society*, 15(6), pp. 853–871. Available at: https://doi.org/10.1177/1461444812463345.

Parasie, S. (2022) *Computing the News: Data Journalism and the Search for Objectivity*. New York, NY: Columbia University Press.

Pariser, E. (2011) *The Filter Bubble: What the Internet is Hiding from You*. London: Penguin.

Paulussen, S. (2012) 'Technology and the transformation of news work: are labor conditions in (online) journalism changing?' in E. Siapera and A. Veglis (eds) *The Handbook of*

Global Online Journalism. Hoboken, NJ: Wiley, pp. 192–208. Available at: https://doi.org/10.1002/9781118313978.ch11.

Pavlik, J. (2000) 'The impact of technology on journalism', *Journalism Studies*, 1(2), pp. 229–237. Available at: https://doi.org/10.1080/14616700050028226.

Petre, C. (2021) *All the News That's Fit to Click: How Metrics Are Transforming the Work of Journalists*. Princeton, NJ: Princeton University Press.

Phillips, A. (2010) 'Transparency and the new ethics of journalism', *Journalism Practice*, 4(3), pp. 373–382. Available at: https://doi.org/10.1080/17512781003642972.

Piechota, G. (2020) 'How Helsingin Sanomat automated its COVID-19 updates', *Readers First Initiative Blog*, 25 May. Available at: https://www.inma.org/blogs/reader-revenue/post.cfm/how-helsingin-sanomat-automated-its-covid-19-updates (Accessed: 12 August 2022).

Pignard-Cheynel, N. and Amigo, L. (2019) 'Le chargé des réseaux socio-numériques au sein des médias: entre logiques gatekeeping, marketing et participative', *Réseaux*, 213, pp. 139–172. Available at: https://doi.org/10.3917/res.213.0139.

Plattner, T. and Orel, D. (2019) 'Addressing micro-audiences at scale', *Computation + Journalism Symposium 2019*, Miami, FL, 1–2 February. Available at: http://cplusj.org/program (Accessed: July 29 2021).

Plavec, J. G. (2017) 'Feinstaubradar Stuttgart: so viel feinstaub ist an ihrem ort in der luft', *Stuttgarter Zeitung*, 2 November. Available at: https://www.stuttgarter-zeitung.de/inhalt.feinstaubradar-stuttgart-so-viel-feinstaub-ist-an-ihrem-ort-in-der-luft.91348ea0-3c1d-48b0-abbb-7bd8d399c223.html (Accessed: 14 August 2022).

Plesner, U. (2009) 'An actor-network perspective on changing work practices: communication technologies as actants in newswork', *Journalism*, 10(5), pp. 604–626. Available at: https://doi.org/10.1177/1464884909106535.

Porlezza, C. and Splendore, S. (2019) 'From open journalism to closed data: data journalism in Italy', *Digital Journalism*, 7(9), pp. 1230–1252. Available at: https://doi.org/10.1080/21670811.2019.1657778.

Powers, M. (2012) '"In forms that are familiar and yet-to-be invented": American journalism and the discourse of technologically specific work', *Journal of Communication Inquiry*, 36(1), pp. 24–43. Available: https://doi.org/10.1177/0196859911426009.

Primo, A. (2019) 'Actor-network theory', in T. P. Vos, F. Hanusch, D. Dimitrakopoulou, M. Geertsema-Sligh and A. Sehl (eds) *International Encyclopedia of Journalism Studies*. Hoboken, NJ: Wiley. Available at: https://doi.org/10.1002/9781118841570.iejs0013.

Prior, N. (2008) 'Putting a glitch in the field: Bourdieu, Actor Network Theory and contemporary music', *Cultural Sociology*, 2(3), pp. 301–319. https://doi.org/10.1177/1749975508095614.

Raley, R. (2013) 'Dataveillance and countervailance' in L. Gitelman (ed.) *Raw Data is an Oxymoron*. Cambridge, MA: MIT Press.

Rämö, M. and Leppänen, L. (2021) 'Using contextual and cross-lingual word embeddings to improve variety in template-based NLG for automated journalism', *Proceedings of the EACL Hackashop on News Media Content Analysis and Automated Report Generation*, online, 19 April. Available at: https://aclanthology.org/2021.hackashop-1.9 (Accessed: 12 September 2022).

Rashidian, N., Brown, P. D., Hansen, E., Bell, E. J. and Albright, J. R. (2019) *Friend and Foe: The Platform Press at the Heart of Journalism*. Tow Center for Digital Journalism report. Available at: https://doi.org/10.7916/d8-15pq-x415.

Rashidian, N., Tsiveriotis, G., Brown, P. D., Bell, E. J. and Hartstone, A. (2020) *Platforms and Publishers: The End of an Era*. Tow Center for Digital Journalism report. Available at: https://doi.org/10.7916/d8-sc1s-2j58.

Rédaction du Monde.fr. (2015) 'Des robots au « Monde » pendant les élections départementales ? Oui… et non' *Le Monde,* 23 March. Available at: https://www.lemonde.fr/le-monde/article/2015/03/23/des-robots-au-monde-pendant-les-elections-departementales-oui-et-non_5995670_4586753.html (Accessed: July 29 2021).

Redazione ANSA (2020) 'Coronavirus: Ansa e Applied XLab producono notizie sull'epidemia grazie all'intelligenza artificiale', *ANSA,* 27 April. Available at: https://www.ansa.it/sito/notizie/cronaca/2020/04/27/coronavirus-ansa-e-applied-xlab-producono-notizie-sullepidemia-grazie-allintelligenza-artificiale_7fecc4c3-8c58-4cbe-815c-a46b51684cd4.html (Accessed: 12 August 2022).

Reese, S. D. and Ballinger, J. (2001) 'The roots of a sociology of news: remembering Mr. Gates and Social Control in the Newsroom' *Journalism & Mass Communication Quarterly*, 78(4), pp. 641–658. Available at: https://doi.org/10.1177/107769900107800402.

Regan Shade, L. (2003) 'Technological determinism', in S. Jones (ed.) *Encyclopedia of New Media*. Thousand Oaks, CA: SAGE, pp. 433–434. Available at: https://doi.org/10.4135/9781412950657.n226.

Reiter, E. and Dale, R. (2000) *Building Natural Language Generation Systems*. Cambridge: Cambridge University Press.

Rojas Torrijos, J. L. (2019) 'Automated sports coverages: case study of bot released by The Washington Post during Río 2016 and Pyeongchang 2018 Olympics'. *Revista Latina de Comunicación Social,* 74, pp. 1729–1747. Available at: https://doi.org/10.4185/RLCS-2019-1407en.

Romele, A. (2021) 'Technological Capital: Bourdieu, Postphenomenology, and the Philosophy of Technology Beyond the Empirical Turn', *Philosophy & Technology* 34(2021), pp. 483–505. Available at: https://doi.org/10.1007/s13347-020-00398-4.

Ryfe, D. M. (2018) 'A practice approach to the study of news production', *Journalism*, 19(2), pp. 217–233. Available at: https://doi.org/10.1177/1464884917699854.

Ryfe, D. M. (2022) 'Actor-network theory and digital journalism', *Digital Journalism*, 10(2), pp. 267–283. Available at: https://doi.org/10.1080/21670811.2021.1945937.

Sayes, E. (2014) 'Actor–network theory and methodology: Just what does it mean to say that nonhumans have agency?' *Social Studies of Science*, 44(1), pp. 134–149. Available at https://doi.org/10.1177/0306312713511867.

Schapals, A. K. and Porlezza, C. (2020) 'Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions', *Media and Communication*, 8(3), pp. 16–26. Available at: https://doi.org/10.17645/mac.v8i3.3054.

Scheufele, B. (2008) 'Grounded Theory', in W. Donsbach (ed.), *The International Encyclopedia of Communication*. Hoboken, NJ: Wiley. Available at: https://doi.org/10.1002/9781405186407.wbiecg027.

Schmidt, C. (2019) 'How to cover 11,250 elections at once: here's how The Washington Post's new computational journalism lab will tackle 2020', *Nieman Lab*, 25 July. Available at: https://www.niemanlab.org/2019/07/how-to-cover-11250-elections-at-once-heres-how-the-washington-posts-new-computational-journalism-lab-will-tackle-2020/ (Accessed: 12 August 2022).

Schmitz Weiss, A., and Domingo, D. (2010) 'Innovation processes in online newsrooms as actor-networks and communities of practice', *New Media & Society*, 12(7), pp. 1156–1171. Available at: https://doi.org/10.1177/1461444809360400.

Schneider, C. (2022) 'Neu: Automatisierte Marktberichte im Börsenticker', *Bayerischer Rundfunk*, February 21. Available at: https://www.br.de/nachrichten/wirtschaft/automatisierte-marktberichte-im-boersenticker,SxnmufC (Accessed: 12 August 2022).

Schneider, C. and Köppen, U. (2021) 'Wie Textautomatisierung BR Sport unterstützt', *Bayerischer Rundfunk*, 17 February. Available at: https://www.br.de/nachrichten/sport/wie-textautomatisierung-br-sport-unterstuetzt,SIl2t2b (Accessed: 12 August 2022).

Schudson, M. (1978) *Discovering The News: A Social History Of American Newspapers*. New York: Basic Books.

Schudson, M., 2005. 'Autonomy from what?' In R. Benson and E. Neveu (eds) *Bourdieu and the Journalistic Field*. Cambridge: Polity, pp. 214–223.

Schultz, I. (2007) 'The journalistic gut feeling', *Journalism Practice*, 1(2), pp. 190–207. Available at: https://doi.org/10.1080/17512780701275507.

Schwandt, T. A. (2007) 'Analytic generalization', in T. A. Schwandt (ed.) *The SAGE Dictionary of Qualitative Inquiry*. 3rd Edition. Thousand Oaks, CA: SAGE, pp. 5–6. Available at: https://dx.doi.org/10.4135/9781412986281.n5.

Schwencke, K. (2014) 'Earthquake aftershock: 2.7 quake strikes near Westwood', *The Los Angeles Times*, 17 March. Available at: https://www.latimes.com/local/lanow/earthquake-27-quake-strikes-near-westwood-california-rdivor-story.html (Accessed: July 29 2021).

Scott, C. (2009) 'Grounded Theory', in S. W. Littlejohn and K. A. Foss (eds) *Encyclopedia of Communication Theory: Vol. 1*. Thousand Oaks, CA: SAGE, pp. 448–450. Available at: https://doi.org/10.4135/9781412959384.n166.

Sebis Research News (2021) 'Sebis and Bayerischer Rundfunk launch NLG system for automated basketball news', 18 February. Available at: https://wwwmatthes.in.tum.de/pages/sp7lk5bzaq98/sebis-and-Bayerischer-Rundfunk-launch-NLG-system-for-automated-basketball-news (Accessed: 12 August 2022).

Sehl, A., Cornia, A., Graves, L. and Nielsen, R. K. (2019) 'Newsroom integration as an organizational challenge', *Journalism Studies*, 20(9), pp. 1238–1259. Available at: https://doi.org/10.1080/1461670X.2018.1507684.

Sehl, A. and Cornia, A. (2021) 'How do public service media innovate? An analysis of product development by European PSM', *Journalism Studies*, 22(11), pp. 1469–1486. Available at: https://doi.org/10.1080/1461670X.2021.1937675.

Sehl, A., Cornia, A. and Nielsen, R. K. (2021) 'How do funding models and organizational legacy shape news organizations' social media strategies? A Comparison of Public Service and Private Sector News Media in Six Countries', *Digital Journalism*. Available at: https://doi.org/10.1080/21670811.2021.1968920.

Sehl, A. and Cornia, A. (2021) 'How do public service media innovate? An analysis of product development by European PSM', *Journalism Studies*, 22(11), pp. 1469–1486. Available at: https://doi.org/10.1080/1461670X.2021.1937675.

Siapera, E. and Spyridou, L.-P. (2012) 'The Field of Online Journalism: a Bourdieusian analysis', in E. Siapera and A. Veglis (eds) *The Handbook of Global Online Journalism*. Hoboken, NJ: Wiley, pp. 77–97.

Silver, D. *et al.* (2016) 'Mastering the game of Go with deep neural networks and tree search', *Nature*, 529, pp. 484–489. Available at: https://doi.org/10.1038/nature16961.

Singer, J. B. (2004) 'Strange bedfellows? The diffusion of convergence in four news organizations', *Journalism Studies*, 5(1), pp. 3–18. Available at: https://doi.org/10.1080/1461670032000174701.

Sirén-Heikel, S., Leppänen, L., Lindén, C.-G. and Bäck, A. (2019) 'Unboxing news automation: exploring imagined affordances of automation in news journalism', *Nordic Journal of Media Studies*, 1(1), pp. 47–66. Available at: https://doi.org/10.2478/njms-2019-0004.

Sirén-Heikel, S., Kjellman, M. and Lindén, C.-G. (2022) 'At the crossroads of logics: automating newswork with artificial intelligence—(re)defining journalistic logics from the perspective of technologists', *Journal of the Association for Information Science and Technology*, pp. 1– 13. Available at: https://doi.org/10.1002/asi.24656.

Spyridou, L.-P., Matsiola, M., Veglis, A., Kalliris. G. and Dimoulas, C. (2013) 'Journalism in a state of flux: journalists as agents of technology innovation and emerging news practices', *International Communication Gazette*, 75(1), pp. 76–98. Available at: https://doi.org/10.1177/1748048512461763.

Stavelin, E. (2013) *Computational Journalism: When Journalism Meets Programming*. PhD thesis. University of Bergen. Available at: http://hdl.handle.net/1956/7926 (Accessed: 9 September 2022).

Stefanikova, S. (2019) 'Czechs get to grips with AI-powered journalism', *European Journalism Observatory*, 20 November. Available at: https://en.ejo.ch/digital-news/technology/czechs-get-to-grips-with-ai-powered-journalism (Accessed: January 3 2022).

Størmer Thaulow, S., Næss, I. and Stenbom, A. (2021) 'Introducing the Schibsted FAST framework', *Schibsted*, 29 November. Available at: https://schibsted.com/news/introducing-the-schibsted-fast-framework/ (Accessed: 09 August 2022).

Stray, J. (2019) 'Making artificial intelligence work for investigative journalism', *Digital Journalism*, 7(8), pp. 1076–1097. Available at: https://doi.org/10.1080/21670811.2019.1630289.

Sundar, S. S. (1999) 'Exploring receivers' criteria for perception of print and online news', *Journalism & Mass Communication Quarterly*, 76(2), pp. 373–386. Available at: https://doi.org/10.1177/107769909907600213.

Sundar, S. S. (2008) 'The MAIN Model: a heuristic approach to understanding technology effects on credibility', in M. J. Metzger and A. J. Flanagin (eds) *Digital Media, Youth, and Credibility*. Cambridge, MA: MIT Press, pp. 73–100.

Sundar, S. S. and Nass, C. (2001) 'Conceptualizing sources in online news', *Journal of Communication*, 51(1), pp. 52–72. Available at: https://doi.org/10.1111/j.1460-2466.2001.tb02872.x.

Sunstein, C. R. (2017) *#Republic*. Princeton, NJ: Princeton University Press.

Swartz, A. (2013) 'Aaron Swartz's a programmable web: an unfinished work', in J. Hendler and Y. Ding (eds) *Synthesis Lectures on the Semantic Web: Theory and Technology*. San Rafael, CA: Morgan and Claypool. Available at: http://www.morganclaypoolpublishers.com/catalog_Orig/9781627051699_WEB.pdf (Accessed: 9 September 2022).

Syllabs, (no date) *Case Studies*. Available at: https://www.syllabs.com/en/client/ (Accessed: 20 May, 2022).

Tandoc, E. C. (2018) 'Five ways BuzzFeed is preserving (or transforming) the journalistic field', *Journalism*, 19(2), pp. 200–216. Available at: https://doi.org/10.1177/1464884917691785.

Tandoc, E. C. and Jenkins, J. (2017) 'The buzzfeedication of journalism? How traditional news organizations are talking about a new entrant to the Journalistic Field will surprise you!' *Journalism*, 18(4), pp. 482–500. Available at: https://doi.org/10.1177/1464884915620269.

Tandoc, E. C., and Foo, C. Y. W. (2018) 'Here's what BuzzFeed journalists think of their journalism', *Digital Journalism*, 6(1), pp. 41–57. Available at: https://doi.org/10.1080/21670811.2017.1332956.

Tandoc, E. C. and Maitra, J. (2018) 'News organizations' use of native videos on Facebook: tweaking the journalistic field one algorithm change at a time', *New Media & Society*, 20(5), pp. 1679–1696. Available at: https://doi.org/10.1177/1461444817702398.

Tandoc, E. C., Lim, J. Y., Wu, S. (2020) 'Man vs. Machine? The impact of algorithm authorship on news credibility', *Digital Journalism*, 8(4), pp. 548–562. Available at: https://doi.org/10.1080/21670811.2020.1762102.

Thurman, N. (2008) 'Forums for citizen journalists? Adoption of user generated content initiatives by online news media', *New Media & Society*, 10(1), pp. 139–157. Available at: https://doi.org/10.1177/1461444807085325.

Thurman, N., Dörr, K. N. and Kunert, J. (2017) 'When reporters get hands-on with robo-writing', *Digital Journalism*, 5(10), pp. 1240–1259. Available at: https://doi.org/10.1080/21670811.2017.1289819.

Thurman, N., Stares, S. and Koliska, M. (2022) 'Audience evaluations of news videos made with various levels of automation: a population-based survey experiment' [Abstract]. 72nd Annual ICA Conference, Paris, 26–30 May.

Toporoff, S. (2017) 'The air you breathe in Europe's car capital' [Medium], 7 December. Available at: https://medium.com/editors-lab-impact/the-air-you-breathe-in-europes-car-capital-88cde48da5f9 (Accessed: 14 August 2022).

Tuchman, G. (1973) 'Making news by doing work: routinizing the unexpected', American Journal of Sociology, 79(1), pp. 110–131. Available at: https://www.jstor.org/stable/2776714 (Accessed: 9 September 2022).

Tuchman, G. (1978) *Making News: A Study in the Construction of Reality*. New York, NY: Free Press.

Túñez-Lopez, J. M., Toural-Bran, C., Valdiviezo-Abad, C. (2019) 'Automation, bots and algorithms in newsmaking: impact and quality of artificial journalism', *Revista Latina de Comunicación Social*, 74, pp. 1411–1433. Available at: https://doi.org/10.4185/RLCS-2019-1391en.

United Robots (2021) *How Omni provides crucial coverage with Corona Watch*, 17 November. Available at: https://youtu.be/ScHmfK3gNP4 (Accessed: 12 August 2022).

United Robots, (no date) *Now offering the Corona Watch—part of our Breaking Desk Solution*. Available at: https://www.unitedrobots.ai/about/news/now-offering-the-corona-watch-part-of-our-breaking-desk-solution (Accessed: 12 August 2022).

Usher, N. (2014) *Making News at the New York Times*. Ann Arbor, MI: University of Michigan Press.

Usher, N. (2017) 'Venture-backed news startups and the field of journalism', *Digital Journalism*, 5(9), pp. 1116–1133. Available at: https://doi.org/10.1080/21670811.2016.1272064.

Usher, N. (2018) 'Hacks, hackers, and the expansive boundaries of journalism', in S. A. Eldridge II and B. Franklin *The Routledge Handbook of Developments in Digital Journalism Studies*. Abingdon-on-Thames: Routledge, pp. 348–359.

Usher, N. (2019) 'Hacks and hackers', in T.P. Vos, F. Hanusch, D. Dimitrakopoulou, M. Geertsema-Sligh and A. Sehl (eds) *The International Encyclopedia of Journalism Studies*. Hoboken, NJ: Wiley. Available at: https://doi.org/10.1002/9781118841570.iejs0254.

van Dalen, A. (2012) 'The algorithms behind the headlines', *Journalism Practice*, 6(5–6): pp. 648–658. Available at: https://doi.org/10.1080/17512786.2012.667268.

Van den Bulck, H. and Moe, H. (2018) 'Public service media, universality and personalisation through algorithms: mapping strategies and exploring dilemmas', *Media, Culture & Society*, 40(6), pp. 875–892. Available at: https://doi.org/10.1177/0163443717734407.

van der Kaa, H. and Krahmer, E. (2014) 'Journalist versus news consumer: the perceived credibility of machine written news' *Computation + Journalism Symposium 2014*, New York, NY, 24–25 October. Available at: https://research.tilburguniversity.edu/en/publications/journalist-versus-news-consumer-the-perceived-credibility-of-mach (Accessed: 9 September 2022).

van Dijck, J. (2014) 'Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology' *Surveillance and Society*, 12(2), pp. 197–208. Available at: https://doi.org/10.24908/ss.v12i2.4776.

van Dijck, J., Poell, T. and de Waal, M. (2018). *The Platform Society : Public Values in a Connective World*. Oxford: Oxford University Press.

Vandenberghe, F. and Peters, G. (2021) 'Bourdieu, Pierre (1930–2002)', in G. Ritzer (ed.) *The Blackwell Encyclopedia of Sociology*. Hoboken, NJ: Wiley-Blackwell. Available at: https://doi.org/10.1002/9781405165518.wbeosb043.pub2.

Visvam Devadoss, A. K. V., Thirulokachander, V. R. and Devadoss, A. K. (2019) 'Efficient daily news platform generation using natural language processing', *International Journal of Information Technology*, 11, pp. 295–311. Available at: https://doi.org/10.1007/s41870-018-0239-4.

Volti, R. (2022) 'Technological determinism', in G. Ritzer (ed.) *The Blackwell Encyclopedia of Sociology*. Hoboken, NJ: Wiley-Blackwell. Available at: https://doi.org/10.1002/9781405165518.wbeost010.pub2.

Vos, T. P., Craft, S. and Ashley, S. (2012) 'New media, old criticism: bloggers' press criticism and the Journalistic Field', *Journalism*, 13(7), pp. 850–868. Available at: https://doi.org/10.1177/1464884911421705.

Vos, T. P. (2019) 'Field Theory and journalistic capital', in T. P. Vos, F. Hanusch, D. Dimitrakopoulou, M. Geertsema-Sligh and A. Sehl (eds) *The International Encyclopedia of Journalism Studies*. Hoboken, NJ: Wiley. Available at https://doi.org/10.1002/9781118841570.iejs0020.

Waddell, T. F. (2018) 'A robot wrote this? How perceived machine authorship affects news credibility', *Digital Journalism*, 6(2), pp. 236–255. Available at: https://doi.org/10.1080/21670811.2017.1384319.

Waddell, T. F. (2019a) '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*, 13(10), pp. 1255–1272. Available at: https://doi.org/10.1080/17512786.2019.1585197.

Waddell, T. F. (2019b) 'Can an algorithm reduce the perceived bias of news? Testing the effect of machine attribution on news readers' evaluations of bias, anthropomorphism, and credibility', *Journalism & Mass Communication Quarterly*, 96(1), pp. 82–100. Available at: https://doi.org/10.1177/1077699018815891.

Walker Guevara, M. (2019) 'We used AI to identify the sex of 340,000 people harmed by medical devices', *International Consortium of Investigative Journalists*, 25 November. Available at: https://www.icij.org/investigations/implant-files/we-used-ai-to-identify-the-sex-of-340000-people-harmed-by-medical-devices/ (Accessed: July 29 2021).

Wang, Z., She, Q. and Ward, T. E. (2020) 'Generative adversarial networks in computer vision: a survey and taxonomy', *arXiv*. Available at: https://arxiv.org/abs/1906.01529 (Accessed: August 23 2022).

WashPost PR Blog (2016a) 'The Washington Post experiments with automated storytelling to help power 2016 Rio Olympics coverage', *The Washington Post*, 5 August. Available at: https://www.washingtonpost.com/pr/wp/2016/08/05/the-washington-post-experiments-with-automated-storytelling-to-help-power-2016-rio-olympics-coverage/ (Accessed: 12 August 2022).

WashPost PR Blog (2016b) 'The Washington Post to use artificial intelligence to cover nearly 500 races on Election Day', *The Washington Post*, 19 October. Available at : https://www.washingtonpost.com/pr/wp/2016/10/19/the-washington-post-uses-artificial-intelligence-to-cover-nearly-500-races-on-election-day/ (Accessed: 12 August 2022).

WashPost PR Blog (2017) 'The Washington Post leverages automated storytelling to cover high school football', *The Washington Post*, 1 September. Available at: https://www.washingtonpost.com/pr/wp/2017/09/01/the-washington-post-leverages-heliograf-to-cover-high-school-football/ (Accessed:12 August 2022).

WashPost PR Blog (2020a) 'The Washington Post unveils new data-driven strategy for 2020 coverage', *The Washington Post*, 3 February. Available at: https://www.washingtonpost.com/pr/2020/02/03/washington-post-unveils-new-data-driven-strategy-2020-coverage/ (Accessed: 14 August 2022).

WashPost PR Blog (2020b) 'The Washington Post to debut AI-powered audio updates for 2020 election results', *The Washington Post*, 13 October. Available at: https://www.washingtonpost.com/pr/2020/10/13/washington-post-debut-ai-powered-audio-updates-2020-election-results/ (Accessed: 14 August 2022).

Waterloo News (2019) *New tool uses AI to flag fake news for media fact-checkers*, 16 December. Available at: https://uwaterloo.ca/news/news/new-tool-uses-ai-flag-fake-news-media-fact-checkers (Accessed: July 29 2021).

Watts, R. and Joiner, S. (2020) 'Automating our coronavirus coverage' [Medium], 28 April. Available at: https://medium.com/digital-times/how-we-automated-our-coronavirus-coverage-ca9d6ad8ab8f (Accessed: 16 August 2022).

Weber, M. (1976) 'Towards a sociology of the press', *Journal of Communication*, 26(3), pp. 96–101. Available at: https://doi.org/10.1111/j.1460-2466.1976.tb01910.x.

Weber, M. (1998) 'Preliminary report on a proposed survey for a sociology of the press', *History of the Human Sciences*, 11(2), pp. 111–120. Available at: https://doi.org/10.1177/095269519801100207.

Weeks, L. (2014) 'Media law and copyright implications of automated journalism', *Journal of Intellectual Property and Entertainment Law*, 4(1), pp. 67–94. Available at: https://jipel.law.nyu.edu/vol-4-no-1-3-weeks/ (Accessed: 9 September 2022).

West, B. (2021) 'Dedifferentiation', in G. Ritzer (ed.) *The Blackwell Encyclopedia of Sociology*. Hoboken, NJ: Wiley-Blackwell. Available at: https://doi.org/10.1002/9781405165518.wbeos1393.

Westergaard, J. (1977) 'Power, Class and the Media', in J. Curran, M. Gurevitch, and J. Woollacott (eds) *Mass Communication and Society*. Thousand Oaks, CA: SAGE, pp. 15–115.

Westlund, O. (2011) *Cross-media News Work - Sensemaking of the Mobile Media (R)evolution*. PhD thesis. University of Gothenburg. Available at: http://hdl.handle.net/2077/28118 (Accessed: 9 September 2022).

Wheatley, D. (2020) 'A typology of news sourcing: routine and non-routine channels of production', *Journalism Practice*, 14(3), pp. 277–298. Available at: https://doi.org/10.1080/17512786.2019.1617042.

Willens, M. (2019) 'Forbes is building more AI tools for its reporters', *Digiday*, 3 January. Available at: https://digiday.com/media/forbes-built-a-robot-to-pre-write-articles-for-its-contributors/ (Accessed: July 29 2021).

Willis, P. and Trondman, M. (2000) 'Manifesto for Ethnography', *Ethnography*, 1(1), pp. 5–16. Available at: https://doi.org/10.1177/14661380022230679.

Wing, J. M. (2008), 'Computational thinking and thinking about computing', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366, pp. 3717–3725. Available at: https://doi.org/10.1098/rsta.2008.0118.

Winthereik, B. R. (2020) 'Is ANT's radical empiricism ethnographic?' in A. Blok, I. Farías and C. Roberts (eds) *The Routledge Companion to Actor-Network Theory*. Abingdon-on-Thames: Routledge, pp. 24–33.

Wölker, A. and Powell, T. E. (2021) 'Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism' Journalism, 22(1), pp. 86–103. https://doi.org/10.1177/1464884918757072.

Woodman, S. (2019) 'Using the power of machines to complete impossible reporting tasks', *International Consortium of Investigative Journalists*, 22 October. Available at: https://www.icij.org/inside-icij/2019/10/using-the-power-of-machines-to-complete-impossible-reporting-tasks/ (Accessed: July 29 2021).

Wu, S., Tandoc, E. C. and Salmon, C. T. (2019a) 'Journalism reconfigured: assessing human–machine relations and the autonomous power of automation in news production', *Journalism Studies*, 20(10), pp. 1440–1457. Available at: https://doi.org/10.1080/1461670X.2018.1521299.

Wu, S., Tandoc, E. C. and Salmon, C. T. (2019b) 'A Field analysis of journalism in the Automation Age: understanding journalistic transformations and struggles through structure and agency', *Digital Journalism*, 7(4), pp. 428–446. Available at: https://doi.org/10.1080/21670811.2019.1620112.

Wu, S., Tandoc, E. C. and Salmon, C. T. (2019c) 'When journalism and automation intersect: assessing the influence of the technological field on contemporary newsrooms', *Journalism Practice*, 13(10), pp. 1238–1254. Available at: https://doi.org/10.1080/17512786.2019.1585198.

Wu, Y. (2020) 'Is automated journalistic writing less biased? An experimental test of auto-written and human-written news stories', *Journalism Practice*, 14(8), pp. 1008–1028. Available at: https://doi.org/10.1080/17512786.2019.1682940.

Xiao, H. (2010) 'Single-case designs', in A. J. Mills., G. Durepos and E. Wiebe (eds). Encyclopedia of Case Study Research. Thousand Oaks, CA: SAGE, pp. 868–870. Available at: https://doi.org/10.4135/9781412957397.

Xuequan, M. (2018) 'Xinhua to integrate AI into news production', *XinhuaNet*, 9 January. Available at: http://www.xinhuanet.com/english/2018-01/09/c_129786724.htm (Accessed: 9 September 2022).

Yin, R. K. (1994) *Case Study Research: Design and Methods*. 2nd edition. Thousand Oaks, CA: SAGE.

Young, M. L., Hermida, A. (2015) 'From Mr. and Mrs. Outlier to Central Tendencies: computational journalism and crime reporting at the Los Angeles Times', *Digital Journalism*, 3(3), pp. 381–397. Available at: https://doi.org/10.1080/21670811.2014.976409.

Zamith, R. and Braun, J. A. (2019) 'Technology and journalism', in T. P. Vos, F. Hanusch, D. Dimitrakopoulou, M. Geertsema-Sligh and A. Sehl (eds) *The International Encyclopedia of Journalism Studies*. Hoboken, NJ: Wiley. Available at: https://doi.org/10.1002/9781118841570.iejs0040.

Zeisler, A. and Schmidt, K. (2021) 'Using R&D strategically in WSJ markets' [Medium], 8 March. Available at: https://medium.com/the-wall-street-journal/using-r-d-strategically-in-wsj-markets-a9b6b1b01024 (Accessed: July 29 2021).

Zelizer, B. (1995) 'Journalism's "last" stand: wirephoto and the discourse of resistance', *Journal of Communication*, 45(2), pp. 78–92. Available at: https://doi.org/10.1111/j.1460-2466.1995.tb00729.x.

Zheng, Y., Zhong, B. and Yang, F. (2018) 'When algorithms meet journalism: the user perception to automated news in a cross-cultural context', *Computers in Human Behavior*, 86, pp. 266–275. Available at: https://doi.org/10.1016/j.chb.2018.04.046.

Zuboff, S. (2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.