

# The datafication of digital journalism: A history of everlasting challenges between ethical issues and regulation

Journalism  
2024, Vol. 25(5) 1167–1185  
© The Author(s) 2023



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/14648849231190232

[journals.sagepub.com/home/jou](https://journals.sagepub.com/home/jou)



**Colin Porlezza** 

Università della Svizzera italiana, Lugano, Switzerland

Department of Journalism, School of Communication and Creativity, City, University of London, London, England

## Abstract

Data permeates nearly all spheres of society, and journalism is no exception to this since data has become a cornerstone of reality construction and perception. This contribution sets out to historicize the datafication processes in digital journalism and the way in which European institutions of media (self-)regulation have dealt with ethical issues regarding the use of data in algorithmic journalism in three areas: accountability, transparency, and privacy. The article shows that the process of datafication in journalism cannot be observed and analyzed in isolation, given that there is a double reflexivity between data-driven societal transformation processes and what happens in journalism. However, almost all press councils in Europe have so far ignored data-driven phenomena like algorithms or news automation. As a consequence, if self-regulators do not regulate, other institutions will, with the risk of news organizations being forced to make decisions on the grounds of regulatory frameworks that are not primarily intended for journalism.

## Keywords

Journalism, datafication, algorithms, automation, artificial intelligence, big data, self-regulation, ethics

---

## Corresponding author:

Colin Porlezza, Faculty of Communication, Culture and Society, IMeG – Institute of Media and Journalism, USI Università della Svizzera italiana, Via Buffi 13, Lugano, 6900, Switzerland.

Email: [colin.porlezza@usi.ch](mailto:colin.porlezza@usi.ch)

## Introduction

In 1967, technology theorist Lewis Mumford coined the concept of the “megamachine” in his book *The Myth of the Machine: Technics and Human Development*, situating it in ancient Egypt civilization in relation to the construction of the great pyramids. According to Mumford, these “megamachines” are very large groups of enslaved humans used as “labor devices” and controlled by the power of god-like kings (1967). The cogs in this labor-megamachine are not made of technology, but of human beings that were amassed to carry out the ceaseless work on the pyramids. In his later work *The Myth of the Machine: The Pentagon of Power*, Mumford adapted his concept of the “megamachine” to the digital realm, stating that the Egyptian Sun-God has been replaced by a so called “Omni Computer”, that is “the latest model I.B.M. computer, zealously programmed by Dr. Strangelove and his associates” (1970, p. 273). There is no longer an army of thousands of coerced human beings working together to build pyramids, but rather a power-monopolizing “info-machine” ready to elaborate “quantitatively measured or objectively observed” data. In his critical vision of digitalization, Mumford prophesized our increasing dependency on machines, which would lead to “a total destruction of autonomy”, in which human beings become zombie-like “button-pressers” to keep the machines that improve our lives working. Trying to maintain the machines working, humans themselves are reduced to a cog in the machine, becoming “both producer and product of the new megamachine age” (1970, p. 275).

What Mumford described in the early 1970s is nowadays critically reflected by the increasing datafication of digital journalism: the use of data, algorithms, and machine learning, as well as the technological infrastructures they depend on, represent a new kind of “info-machine” that defines to a large extent how news is gathered, produced, and disseminated (Pavlik, 2013). Overall, as Lewis and Westlund (2015, p. 449) argued, “big data embodies emerging ideas about, activities for, and norms connected with data sets, algorithms, computational methods, and related processes and perspectives tied to quantification as a key paradigm of information work.” Data becomes a main ingredient in the way journalism can be optimized, evaluated, curated, or made more efficient. The sociologist Jason Sadowski (2019) argues in this regard: “The supposed universality of data reframes everything as falling under the domain of data capitalism. All spaces must be subjected to datafication.” Datafication is therefore the foundational layer on which the current transformation of journalism is built.

The concept of datafication was introduced by Mayer-Schönberger and Cukier in their 2013 book *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. The two authors describe datafication as the collection, quantification, and analysis of large quantities of information, turning it into a scalable resource both for knowledge production and the generation of economic value. The data deluge that comes with digitalization entails a growing collection and processing of data that has since “become pervasive and unprecedented in scope and scale, and increasingly automated” (Flensburg and Lomborg, 2021: p. 2).

In journalism, data has a more troubled history, given the long-established skepticism of journalists regarding (technological) change. Ryfe (2012) demonstrated how initial

plans to establish initial digital strategies were met with open resistance. Eventually, digitalization allowed for a computational turn (Coddington, 2015), and boosted a quantitatively oriented journalism (Splendore, 2016). At the same time, data started to influence newsrooms in the form of metrics, affecting in turn the use and allocation of resources, production workflows, the selection and the placement of topics, as well as the format and the style of news (Christin, 2020; Fürst, 2020). With the rise of artificial intelligence, machine learning and algorithms (Kotenidis and Veglis, 2021), the use of data became even more pervasive in journalism as news organizations started to automate news production and distribution, trying to make news work more efficient (Beckett, 2019).

The digitalization of journalism bears specific ethical issues. However, institutions of self-regulation such as press councils tend to slowly adapt to data-driven transformations in journalism: phenomena like personal data, algorithms, or artificial intelligence largely remain “uncharted territory” (Porlezza and Eberwein, 2022). The goal of this article is thus to historicize the datafication processes of digital journalism in Europe, particularly regard the pervasive use of algorithms in newsrooms, and to analyze the way in which European institutions of media (self-) regulation deal with the ethical issues the emerge with algorithmic journalism in three areas: accountability, transparency, and privacy. The article has thus two aims: first, the contribution wants to critically evaluate the significance of datafication in journalism from a historic perspective. Secondly, the article wants to critically discuss the ethical implications of algorithmic journalism and how media self-regulators react to it through codified values and norms. Drawing on a historical approach, the article tries therefore to answer the following two research questions:

**RQ1:** To what extent has datafication become a corner stone of journalistic practice and knowledge production, in particular since algorithms have entered the newsrooms?

**RQ2:** To what extent are institutions of self-regulation regulating the use of data in algorithmic journalism in terms of transparency, accountability, and privacy?

The article proceeds in four sections: first, the paper offers a historic perspective of the datafication of media and society, after that it turns to the impact of data on journalism. Then, the article analyzes the current situation regarding journalism’s use of algorithms and automation. In the last section, the article offers a qualitative analysis of how self-regulators in Europe tackle (or not) ethical concerns related to the use of data in terms of accountability, transparency, and privacy, before offering some concluding remarks.

## The datafication of media and society

Although datafication was introduced into the academic discourse by Mayer-Schönberger and Cukier (2013), the concept is all but new: it has a long history when it comes to population management and bureaucracy in states and organizations (Porter, 1995). Today we relate datafication mostly to digital technology and big data. Datafication has

always been assisted by media technologies, although, in the earlier days, media were analogue and supported by writing, printing, archiving and indexing (Mayer-Schönberger and Cukier, 2013). In this regard, datafication precedes digitalization (Flensburg and Lomborg, 2021: p. 2). Still, since the end of the 1980s, digital infrastructures and the digitalization through the Web have massively increased the generation, processing, and analysis of data, both of private companies and public institutions (Andrejevic, 2014). (Kitchin 2014: p. 80) showed that the rapid growth of data in society is due to the simultaneous development of different technologies, infrastructures, techniques and processes, but also their implementation and embedding in social practices, spaces, and businesses:

“This new knowledge infrastructure includes the widespread roll-out of a diverse set of information and communication technologies, especially fixed and mobile Internet; the embedding of software into all kinds of objects, machines and systems, transforming them from ‘dumb’ to ‘smart’ as well as the creation of purely digital devices and systems; the development of ubiquitous computing and the ability to access networks and computation in many environments and on the move; advances in database design and systems of information management; distributed and forever storage of data at affordable costs; and new forms of data analytics designed to cope with data abundance as opposed to data scarcity.”

The pervasiveness of personal and data-producing technologies such as mobile phones contributes to the creation of data-driven societies. This mobile ecosystem (Pybus and Coté, 2021), combined with the fact that data permeate nearly all spheres of society (Koenen et al., 2021), ensures that all of us contribute to a tight assemblage of data-driven practices that range from platforms controlling the users’ data flow to feed their business models, to the creation of forms of surveillance capitalism (Lyon, 2019; van Dijck et al., 2019). Since the late 2000s, digital intermediaries like social media platforms have boosted data collection through automation, inducing a “significant shift in governance, in which big data analysis is used to predict, preempt, explain and respond to a range of social issues” (Dencik, 2019, p. x). These Silicon Valley corporations have specialized in a form of “extractivism” through the appropriation of data about our lives (Couldry and Mejias, 2020). Data has thus become a determining element in almost all spheres of our lives, and together with the use of algorithms they enable new cultural and social forms (Uricchio, 2017: p. 126).

Early research on datafication offered a general critique of the increasing dataism and blind belief in the power of big data (boyd & Crawford, 2012; Van Dijck, 2014). However, as (Couldry 2004: p. 117) suggested, data and algorithms and their effects on specific practices need to be “decentered”. We therefore need to put the datafication of journalism in a larger context of societal transformations and not observe it in isolation. First, because technology is not a transformational force in itself, but gets implemented grounded on existing value systems that have cultural, social and economic roots (Örnebring, 2010: p. 68). Anderson (2015) also notes that the “technological, social, and cultural influences, and the genealogy of journalistic evidence helps us situate the trend of datafication in journalism into these historical contingencies.” Secondly, because there are mutual

dependencies when it comes to journalism's relation with data: while the overall datafication of society contributed to making more data sources available to journalism, the wide use of algorithms in society has also become a topic of interest for journalism. Datafication thus entails a reflexivity between the ongoing datafication of society, and the datafication of the journalistic field itself: the "algorithmic turn" (Napoli, 2014) in journalism, which involves a growing use of data, algorithms, and machine learning in news work in order to analyze the datafication of society, relies

"on the same means – data and algorithms – that distinguish the datafied society itself. This entails a reflexivity between the instruments that characterize a datafied society, and their implementation in journalism in order to observe them." (Porlezza, 2018, p. 369)

Loosen (2018, p. 17) supports this perspective, asserting that "datafication goes far beyond journalism: it affects the very nature of society's communicative foundations." The impact of the increased use of data practices in journalism - in particular by observing the widespread use of data in society - extends beyond its boundaries and contributes therefore to the datafication of the public sphere itself, with specific ethical challenges for audiences e.g. in terms of privacy, discrimination, and (economic) exploitation.

## The datafication of journalism

Datafication has consequences for journalism, although in distinctive ways at different times. In the 17<sup>th</sup> century, periodicals that offered news from distance cities were data-rich because they carried information about business details (Klein, 2015). Towards the end of the 18<sup>th</sup> century, newspapers such as the Wall Street Journal started to use statistical data and tables to report about wars, mortality rates, or public affairs (Usher, 2016). Similarly, sports journalism in the U.S. evolved to a more data-centered practice, as journalists in the late 19<sup>th</sup> century started to record data of baseball games, to analyze them statistically and to present them in box scores (Schwarz, 2004).

After an initial fascination for data, Anderson, 2018 showed that journalists in the U.S. lost their close relation with data during the "Progressive Era" in the early 20<sup>th</sup> century, for they distanced themselves from social sciences and sociologists after the 1920s in an attempt of boundary work. This approach was also reflected by the opposition of journalists towards sociology's focus on "social structures and depersonalized contextual information, preferring to retain their individualistic focus on powerful personalities and important events" (Anderson, 2020).

Until the 1950s, data and data-driven infrastructures play a subordinate role. But in 1952 a Remington Rand UNIVAC computer was used to predict the victory of Dwight D. Eisenhower in the U.S. presidential election - launching therefore the era of computer-assisted-reporting (CAR), which is often seen as the moment of birth of data journalism. But only in combination with Phil Meyer's understanding of a more social scientific and research methods-oriented "precision journalism" in the 1960s we can understand the new foundation of what would later become data journalism. These ruptures led to a more quantitatively oriented journalism, focusing on structures rather than just on episodic

events, took place in the 1960s and 1970s and not only after the technological innovations of the Web in the early 1990s (Anderson, 2015).

Early research into data-driven journalism revolved around questions regarding the transformation of the journalistic profession, organizational changes, as well as the specific skills data journalists have compared to other journalists working in the same newsroom (Ausserhofer et al., 2020). But only with the rise of data journalism as a “new” genre (Lewis, 2015: p. 322-323), data also became more present in journalism scholarship. This contributed - together with the rapid technological developments in the newsrooms - to the proliferation of different labels, ranging from “data-driven journalism” (Parasie and Dagiral, 2013), “interactive journalism” (Usher, 2016), to “big data journalism” (Tandoc and Oh, 2017). Of particular importance for the European perspective was the adoption of “data journalism” as a new beat by The Guardian in 2011 through a defining article written by Simon Rogers (2011), which entailed the adoption of the expression both in Europe and in the U.S. (Gynnild, 2014). Additionally, in 2017 the European Data Journalism Network (EDJNet), a network of media organizations across Europe promoting data journalism on European issues, was founded. These events launched an initial spark of specifically European research into the topic of datafication. Before that, research on data in journalism was scattered. McNair (1994) for instance looked into the supply of information through journalism and raw data while talking about the British media system. In 2007, MacGregor published an article about the use of metrics and the way online journalists reacted to new ways of gaining knowledge about their audiences in some European news organizations. But only with the increasing scholarly investigations into data journalism the specific European context of datafied journalism became more visible. Ausserhofer et al. (2020) showed that only from 2010 on there was an increase in research on data and journalism, although CAR existed since the 1960s. Surprisingly, most authors originated from Europe, with investigations into data-intensive newswork in several countries such as e.g. Norway (Karlsen and Stavelin, 2014), the UK (Hannaford, 2015), Sweden (Appelgren and Nygren, 2014), Germany (Weinacht and Spiller, 2014), and Belgium (De Maeyer et al., 2015).

A few scholars have engaged in clarifying various typologies of quantitative forms of journalism (Coddington, 2015; Gynnild, 2014; Splendore, 2016; Usher, 2016). Overall, three forms emerge from the academic literature: data journalism, computational journalism, and automated journalism. Data journalism can be regarded as “a form of storytelling where traditional journalistic working methods are mixed with data analysis, programming, and visualization techniques” (Splendore, 2016).

Compared to data journalism, computational journalism encompasses “finding, telling, and disseminating news stories with, by, or about algorithms” (Diakopoulos and Koliska, 2017: p. 810). Computational journalism has a shorter history, but its scope broadened considerably over the last years (Coddington, 2018; Flew et al., 2012; Thurman, 2019). Early publications concentrated on the differentiation between CAR and computational journalism. In 2011, Diakopoulos wrote that computational journalism was inclusive of CAR, but differed in terms of its processing capabilities. Coddington (2015, p. 336) instead argued that “computational journalism goes beyond CAR in its focus on the processing capabilities of computing, particularly aggregating, automating, and

abstracting information”. Yet, defining computational journalism as the application of technology to journalism does not suffice if one wants to understand the changing relationship between journalism practice, data, and algorithms. According to [Caswell \(2019, p. 1137\)](#), who works for the BBC, computational journalism is “a practice in which journalistic knowledge is represented computationally, as structured data, during reporting, analysis, distribution or consumption. (...) In this view, therefore, computational journalism is fundamentally an interpretation of the nature of journalistic knowledge and its representation.” Computational journalism has thus moved away from the notion of simply using more powerful machines in news work and instead denotes the broad and pervasive application of algorithms in all stages of the news cycle.

The last typology, automated journalism, is originally “conceptualized as algorithmic processes that convert data into narrative news texts with limited to no human intervention beyond the initial programming” ([Carlson, 2015: p. 417](#)). Today, however, automated journalism largely describes the autonomous gathering, production, and distribution of news ([Danzon-Chambaud, 2021](#)), and has therefore replaced computational journalism as a typology. As algorithms - not only computers - are becoming more pervasive, even if mostly in larger news organizations, they “influence, to some extent, nearly every aspect of journalism” ([Zamith, 2019](#)). Automation is being used from news gathering ([Thurman et al. 2017](#)), to news production ([Carlson, 2015, 2018](#); [Diakopoulos, 2019](#); [Dörr, 2016](#)), distribution ([Ford and Hutchinson 2019](#)), and personalization ([Thurman 2019b](#)).

## From data to algorithms and automation

Towards the end of the 2010s, the news media industry became interested in artificial intelligence (AI) since the technology promised to make news work more efficient, more relevant to users, and less expensive ([Beckett, 2019](#)). The hype around AI was also fueled by the promotion of the technology from Silicon Valley, “which often tends to reinforce tendencies whereby managers follow trends or industry hype ([Simon, 2022](#)). This shows yet another element of the datafication process of journalism: it is also about journalism’s own fixation on data that fosters the use as well as the implementation of algorithms and AI technology in news work. As [Meier et al. \(2022\)](#) demonstrate, many of the most important journalistic innovations in the last decade in five European countries (Austria, Germany, Spain, Switzerland, United Kingdom) are linked to data. Among the top 10 spots of innovations in all analyzed countries the researchers identified innovations such as data journalism (the top-ranked innovation), collaborative investigative journalism (ranked 2<sup>nd</sup>), user data to foster engagement (ranked 3<sup>rd</sup>), and automation (ranked 7<sup>th</sup>).

Like any other technological innovation, algorithms not only support journalists in their everyday work, but they impact the nature, role and workflows of journalism ([Thurman et al., 2017, 2019](#)) and contribute to “make journalism in new ways, by creating new genres, practices, and understandings of what news and news work is, and what they ought to be” ([Bucher, 2018: p. 132](#)). In fact, the ubiquity of automated processes exerts a transformational role on the autonomy of journalists ([Van Dalen, 2012](#)), as “the machine is able to work automatically after initial programming to make its own selections of what



content to pull, what figures to populate news stories with, what templates to use, and what content to publish. The human merely plays a ‘checking’ role here” (Wu et al., 2019: p. 1453). Algorithms are therefore increasingly determining editorial decisions, putting journalism’s professional identity in flux, particularly with regard to the question of whether humans remain “in the loop” (Schapals and Porlezza, 2020). However, they entail specific challenges in terms of epistemological claims (Anderson and De Maeyer, 2015) and ethical issues (Lewis and Westlund, 2015).

The increasing interwovenness of humans and machines promise a “hybrid state” of machines and humans collaborating (Milosavljević and Vobič, 2019), a combination that will ultimately lead to what Diakopoulos calls “hybrid journalism” (2019, p. 34). However, it is all but clear “how humans and algorithms [should] be blended together in order to efficiently and effectively produce news information” (Diakopoulos, 2019: p. 8). To make things more complex, automation entails several challenges, not only when it comes to the design of the relation between human judgment and automation (Gutierrez Lopez et al., 2023, Gutierrez-Lopez et al., 2019), but in particular in relation to journalistic values and professional ethics in order to ensure an accountable use of algorithms, in particular because “algorithms are judged, made sense of and explained with reference to existing journalistic values and professional ethics” (Bucher, 2018: p. 129). Not only do algorithms raise ethical issues regarding the objectivity of their output (Steensen, 2019), but also regarding professional journalistic values such as transparency, accountability and responsibility (Porlezza, 2020; Komatsu et al., 2020, p. 3; Dörr and Hollnbuchner, 2017). As a consequence, the introduction of algorithms and automation has often been met with skepticism due to concerns that machines will not be able to act in accordance with journalism ethics (Culver, 2016).

Transparency in particular is a timely issue that is discussed both in scholarly work (Ananny, 2016) and in policy debates.<sup>1</sup> While data journalists cater to norms such as transparency and open-source philosophy (Porlezza, 2019), advocating openness that allows users to access and verify data, this is hardly possible in relation to algorithms. While calls for transparency are often voiced, Ananny and Crawford (2018) demonstrated that the concept is difficult to implement in the field of algorithms due to the complexities related to input-output data, machine learning models, and user interfaces. There are cases, where the code and additional information are made public on GitHub, for instance in the form of readme files that include the specific goal, usage, useful links and other relevant data (Stark and Diakopoulos, 2016).

At the same time, there are several obstacles that make media organizations hesitant about disclosure and a more accountable stance in relation to algorithms. Diakopoulos and Koliska (2017) demonstrate that media companies are afraid of losing competitive advantage by disclosing the code of an algorithm.

Similarly, algorithms pose specific challenges regarding accountability: not only are there risks of biases (Friedman and Nissenbaum, 1996) that can impact data analysis and therefore also news reporting (Gillespie, 2014), but the issue also concerns the way to handle these risks in terms of responsibility. Accountability needs to be implemented at the design level, where processes need to “adjudicate and facilitate the correction of false positives” (Diakopoulos, 2016, p. 58). This also entails transparency as a complementary



principle, allowing users to understand how systems works, and to correct inaccurate data (Diakopoulos, 2016). Even more so as ethical concerns regarding the use of data and algorithms still play a minor role in the news industry, which is also reflected in the inconsistent byline and disclosure policies of news organizations in relation to automated journalism (Montal and Reich, 2017). A responsible approach to the use of algorithmic systems in journalism also comes with a particular AI governance that includes specific checks and balances from the design to the implementation of automated systems (Porlezza, 2023). Taken all this together it becomes difficult to establish an algorithmic accountability that can be evaluated against professional ethical values. The question now is, how is the journalistic profession, that is European institutions of media (self-)regulation, trying to deal with ethical issues regarding the use of data and algorithms in journalism in three areas: accountability, transparency, and privacy?

### **Data, algorithms, and self-regulation**

A previous study (Porlezza and Eberwein, 2022) indicate that institutions of media self-regulation like press councils are not entirely ill-prepared to tackle ethical issues related to digital journalism because ethical codes and guidelines can be applied to the digital field as well. However, most press councils applied a wait and see attitude which reflected in the fact that very few digital-specific regulations were included in the codes: “Albeit the efforts of some stakeholders prove that professional journalistic ethics are in the middle of a process of transformation in the digital age, conclusive sets of guidelines are anything but complete” (Porlezza and Eberwein, 2022: p. 357).

Haapanen (2022) has recently carried out an investigation into European press councils and automation and his findings confirm previous results: press councils are currently waiting and watching. The main reason for their hesitancy can be found in the fact that both news automation and news personalization are still considered to be in their infancy in most media markets - even if automation processes have already become a well-established field within larger news media organizations, particularly in public service media. The AI and Data Group as well as the AI Knowledge Hub at the European Broadcasting Union EBU speak a different language as they represent central topics and areas of innovation. Even if the applications might seem modest, algorithms and automation require data, which is especially true for machine learning. Besides, the fact that current studies confirm a disinterest among large media players for ethical issues (Porlezza and Ferri, 2022), combined with concerns for data, privacy, and transparency should convince self-regulators to become more active in the field.

### *An exploratory study into codes of ethics*

In order to further substantiate the two previous studies, the author carried out an exploratory study into 17 code of ethics of 15 European countries. The analysis was carried out through a document analysis (Prior, 2003) of freely accessible professional codes of ethics<sup>2</sup> in journalism in English, French, Italian or German. In a first desk research, all the relevant documents for the study, that is the code of ethics, were searched, identified, and

selected grounded on a qualitative assessment of the topical reference to ethical guidelines as well as the role of the institution that published the codes of ethics or guidelines. Secondly, the document analysis was combined with a text-mining-approach (Ignatow and Mihalcea, 2018). By using Adobe's Advanced Research, all the documents have been analyzed with the following objective: first, to identify the presence of the four terms "data", "algorithm", "automation", or "artificial intelligence" in the documents. These keywords allowed to identify the relevant paragraphs, rules, or principles within the documents. Secondly, in the case of the identification of any of the four terms in the documents, the author analyzed the way in which the terms are understood in relation to ethical considerations. While the first part produced a quantitative data output that offers insights about how often data-related terms are mentioned in the documents, the second part of the text-mining allowed for a closer qualitative analysis by isolating relevant paragraphs to see what kind of topic the identified elements are about. The analysis builds on previous empirical findings (Porlezza, 2023; Haapanen, 2022) and offers therefore a critical reflection on how ethical issues regarding algorithms and automation are currently tackled (or not) at the professional self-regulatory level.

Haapanen's findings are largely confirmed: when it comes to data, the analyzed code of ethics most often refer to issues of data protection and privacy. For example, the German Press Code refers 24 times to data, and the Preamble states that "The regulations pertaining to editorial data protection apply to the Press in gathering, processing or using information about persons for journalistic-editorial purposes. From research to editing, publishing, documenting and storing these data, the Press must respect people's privacy and right to self-determination on information about them. All person-related data gathered, processed and used for journalistic-editorial purposes are subject to editorial secrecy." However, the code does not mention digital data specifically, and the same goes for most other codes. In the case of the Swiss Press Council, data are mentioned right at the beginning, but on a more general basis referring to the search for truth, which "is at the heart of the act of informing. It presumes taking account of available and accessible data, respect for the integrity of documents (text, recording, image), verification and rectification." Once again, digital data are not mentioned specifically. If digital (or online) data are nominated, only a few press councils like the Catalan one refer to the right to be forgotten: "In the case that they ask for erasure of the data, the requirement should reconcile the public interest with the individual rights."

The Finnish Council for Mass Media is currently one among a few press/media councils that has elaborated a specific statement on news automation and personalization.<sup>3</sup> In 2019, it issued a "Statement on marking news automation and personalization" in which the Finnish Council defined the use of algorithms and automation in relation to journalistic work, but also to define a responsible and transparent use of the technology with respect to the public. The statement focuses on two elements: first, the editorial office should retain the journalistic decision-making power. This strong emphasis not to relinquish decision-making either to a machine or to technologists is in line with research findings that show how journalists, even if they are upbeat about the technology, are quite adamant about carefully automating selected journalistic tasks with tools they feel comfortable about (Gutierrez Lopez et al., 2023). It is ultimately the editor in chief's

responsibility to ensure that the algorithm adheres to journalistic standards regarding its content production. This means that the members of the media outlet need to have sufficient knowledge and “understanding of the effect of algorithmic tools on content. For example, if a media outlet purchases a tool developed externally, the outlet must examine and approve of its central operational principles and be able to react, should problems arise”.<sup>4</sup> This also means that external developers must apply a value-sensitive design strategy, adhering to journalistic guidelines in terms of the output.

Along the lines of [Montal and Reich \(2017\)](#) in terms of byline disclosure, the Finnish Council also states that the public has the right to know about automation and personalization. This request of transparency towards the public is also grounded on its inability to distinguish between content produced by humans and algorithms ([Graefe and Bohlken, 2020](#)). On the same grounds, the Finnish Council also asks news organizations to disclose the use of personalization technology, should the targeted content represent a significant amount of what is offered. However, personalization not only includes ethical issues about transparency, but also regarding accountability because it raises concerns in relation to nudging and filter bubbles. While pushing users into a specific direction through personalized content can improve the diversity of the content to which they are exposed, the filtering of certain content can have detrimental effects to participatory democracy - although sometimes filter-bubbles can also be constructive if they “act as incubators of constructive speech, allowing the more marginalized voices in society to join forces and pluck up the courage to speak out ([Helberger, 2019](#): p. 1004). Privacy, on the other hand, is not mentioned at all as an ethical issue of algorithmic journalism.

Overall, current codes of ethics do not offer any guidance on how data, algorithms, automation, and AI should be used in a transparent, accountable and responsible way. The main reason for that can be traced back to the idea that ethical principles should be applicable to different fields, independently of their actual context. An example of that can be found, once more, in the Declaration of Principles of Professional Journalists in Catalonia:

“The CIC recommends the media to act with special responsibility and exactitude in the case of information or opinions with contents that could provoke discrimination for reasons of gender, ethnicity, belief or social or cultural background; they should avoid in any case the generalizations and the labelling of people because of differential features, whether ethnic, religious, economic or social.”

In this paragraph, all the risks that could stem from the use of automation are included: the fact that news media need to act with special responsibility when it comes to issues of bias and discrimination - challenges that are ubiquitous in the case of algorithmic journalism. But they are not mentioned in relation to algorithmic journalism, because the way they are stated, they can be applied to different areas, from data journalism, to issues of inclusivity in journalistic reporting, to generative AI in the automatic text-production.

## Concluding remarks

This contribution has analyzed the datafication of journalism from a historical perspective. The article showed that the process of datafication in journalism cannot be observed and analyzed in isolation, but needs to take into account the broader context of society. In addition, datafication encompasses a reflexivity between data-driven societal transformation processes and what happens in journalism. In this sense, the algorithmic turn in journalism involves the growing use of data, algorithms, and machine learning in order to analyze the datafication of society. On top of that, there is another reflexive layer due to the fact that the datafication process in journalism re-enforces itself due to the centrality of data: This shows yet another element of the datafication process of journalism when it comes to innovation processes: journalism innovation is largely by data-driven, which in turn fosters the use as well as the impact of data on news work and journalists. The effects of the development can be seen in many different areas within journalism, from changing journalistic role perceptions, identities, to the way humans have to interact and collaborate with machines, or to matters of journalistic autonomy and authority. Hence, to answer RQ1, datafication has become a corner stone of journalistic practice and knowledge production. However, we need to be careful not to overestimate the impact of news automation: fully operational automation of news production is still relatively rare, and human oversight is oftentimes necessary due to fuzzy datasets or unexpected events that the algorithm is unable to elaborate. As [Haapanen \(2022, p. 9\)](#) states, there is “no need to see monsters where they do not exist.” On the contrary, algorithms are often beneficial to journalism because they do help journalists in their investigations, for instance when it comes to the analysis of large data-dumps or leaks.

At the same time, as algorithmic journalism has become a defining feature of news work, ethical (and legal) issues with regard to transparency and accountability are becoming pressing concerns: who takes responsibility for automated journalism? How can we make automated journalism more transparent? How can we make sure that the (personal) data algorithms use are in line with current privacy and data protection laws? News organizations are just starting to be concerned with ethical issues raised by automation ([Porlezza and Ferri, 2022](#)). Although some news organizations have commenced to come up with specific guidelines, many news outlets face specific challenges in defining checks and balances that should ensure a responsible use of AI systems. On top of that, institutions of journalistic self-regulations do not tackle issues related to algorithms in their guidelines (yet), because most institutions either wait and observe, or prefer to develop principles that can be applied to a wide range of issues, not specifically to AI and automation.

To answer RQ2, press councils all over Europe (except for Finland) are hesitant to regulate the use of data in algorithmic journalism. But while press councils adopt a “wait and see” strategy because they perceive news automations as a marginal phenomenon, other institutions have launched full-fledged investigations: both the European Union (AI Act) and the Council of Europe (Guidelines on the responsible use of digital tools including artificial intelligence in journalism) are currently working on the topic. If press councils do not want to regulate, others will. In other words: news organizations might be

forced to make decisions on the grounds of regulatory frameworks that are not primarily intended for journalism. The European Broadcasting Union EBU has already criticized that the risk-based approach taken by the European Commission could threaten the legitimate use of AI systems in the media sector.<sup>5</sup> Hence, the extent to which the use and implementation of data-driven algorithms and automation in journalism becomes even more prominent, a timely discussion of the many ethical issues related to data, privacy, transparency and accountability is necessary and pressing.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iD

Colin Porlezza  <https://orcid.org/0000-0002-1400-5879>

### Notes

1. Besides the current discussion revolving around the European Union's AI Act, see for instance also the work of the Committee of Experts on Freedom of Expression and Digital Technologies with the Council of Europe, or more in general the works of the Council of Europe's Committee on Artificial Intelligence.
2. The analysis included the following codes of ethics: the *Ehrenkodex* for the Austrian Press, the *Code of Practice* of the Belgian Raad voor de Journalistiek, the *Press Ethical Rules* of the Danish Press Council, the *Code of Ethics* for the Estonian Press, the Finnish *Guidelines for Journalists*, the French *Declaration of the Rights and Duties of Journalists*, the German *Press Code*, the Italian *Testo unico dei doveri del giornalista*, the *Guidelines* of the Dutch Raad voor de Journalistiek, the Polish *Media Ethics Charter*, the *Code of Practice* of the Press Council of Ireland, the *Code of Ethics* of the Norwegian Press, the *Código Deontológico* of the Federation of the Press Associations of Spain as well as the *Code of Ethics* of the Information Council of Catalonia (ICC), the Swiss *Declaration of the Duties and Rights of a Journalist* and the related *Directives*, the *Standards* by Impress (UK) as well as the *Editors' Code of Practice* by the Independent Press Standards Organization (IPSO, UK). Only those institutions have been selected that are responsible for the self-regulation of the media or the press.
3. Since January 1, 2019, the Swiss Press Council's jurisdiction also explicitly covers content that has been produced exclusively or with the help of AI. But there are no specific statements or codes that regulate automated news content. However, one of its three Chambers now looks closely into the issue of how to tackle AI. In addition, the Catalan Press Council has also drawn up a series of recommendations for endowing AI with journalistic values (see Ventura [Pocino, 2021](#))

4. “Statement on Marking News Automation and Personalization” <https://jsn.fi/en/lausuma/statement-on-marking-news-automation-and-personalization-2019/>
5. “AI Act: high-risk AI systems need more nuance” <https://www.ebu.ch/news/2022/09/ai-act-high-risk-ai-systems-need-more-nuance>

## References

- Ananny M (2016) Toward an ethics of algorithms convening, observation, probability, and timeliness. *Science, Technology, & Human Values* 41(1): 93–117.
- Ananny M and Crawford K (2018) Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20(3): 973–989.
- Anderson CW (2015) Between the unique and the pattern: historical tensions in our understanding of quantitative journalism. *Digital journalism* 3(3): 349–363.
- Anderson CW (2018) *Apostles of Certainty: Data Journalism and the Politics of Doubt*. Oxford: Oxford University Press.
- Anderson CW (2020) Genealogies of data journalism. In: Grey J and Bounegru L (eds), *The Data Journalism Handbook 2*. <https://datajournalism.com/read/handbook/two/situating-data-journalism/genealogies-of-data-journalism>.
- Anderson C W and De Maeyer J (2015) Objects of journalism and the news. *Journalism* 16(1): 3–9.
- Andrejevic M (2014) Big data, big questions| the big data divide. *International Journal of Communication* 8: 17.
- Appelgren E and Nygren G (2014) Data journalism in Sweden: introducing new methods and genres of journalism into ‘old’ organizations. *Digital Journalism* 2(3): 394–405.
- Ausserhofer J, Gutounig R, Oppermann M, et al. (2020) The datafication of data journalism scholarship: focal points, methods, and research propositions for the investigation of data-intensive newswork. *Journalism* 21(7): 950–973.
- Beckett C (2019) *New Powers, New Responsibilities: A Global Survey of Journalism and Artificial Intelligence*. London: London School of Economics. <https://drive.google.com/file/d/1utmAMCmd4rfJHrUfLLfSJ-clpFTjyefl/view>
- Boyd D and Crawford K (2012) Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society* 15(5): 662–679.
- Bucher T (2018) *If... Then: Algorithmic Power and Politics*. Oxford, UK: Oxford University Press.
- Carlson M (2015) The robotic reporter: automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital journalism* 3(3): 416–431.
- Carlson M (2018) Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism. *New media & society* 20(5): 1755–1772.
- Caswell D (2019) Structured journalism and the semantic units of news. *Digital Journalism* 7(8): 1134–1156.
- Christin A (2020) *Metrics at Work: Journalism and the Contested Meaning of Algorithms*. Princeton: Princeton University Press.
- Coddington M (2015) Clarifying journalism’s quantitative turn: a typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital journalism* 3(3): 331–348.

- Coddington M (2018) Defining and mapping data journalism and computational journalism. In: EldridgeII S and Franklin B (eds), *The Routledge Handbook of Developments in Digital Journalism Studies*. London: Routledge, pp. 225–236.
- Couldry N (2004) Theorising media as practice. *Social Semiotics* 14(2): 115–132.
- Couldry N and Mejias UA (2020) *The Costs of Connection: How Data Are Colonizing Human Life and Appropriating it for Capitalism*. Stanford: Stanford University Press.
- Culver KB (2016) Disengaged ethics. *Journalism Practice* 11(4): 477–492.
- Danzon-Chambaud S (2021) A systematic review of automated journalism scholarship: guidelines and suggestions for future research. *Open Research Europe* 1: 4.
- De Maeyer J, Libert M, Domingo D, et al. (2015) Waiting for data journalism: a qualitative assessment of the anecdotal take-up of data journalism in French-speaking Belgium. *Digital Journalism* 3(3): 432–446.
- Dencik L (2019) Situating practices in datafication—from above and below. In: Stephansen H and Treré E (eds), *Citizen Media and Practice: Currents, Connections, Challenges*. Abingdon and New York: Routledge, pp. 243–255.
- Diakopoulos N (2016) Accountability in Algorithmic Decision Making. *Communications of the ACM* 59(2): 56–62.
- Diakopoulos N (2019) *Automating the News*. Cambridge: Harvard University Press.
- Diakopoulos N and Koliska M (2017) Algorithmic transparency in the news media. *Digital journalism* 5(7): 809–828.
- Dörr K (2016) Mapping the field of algorithmic journalism. *Digital Journalism* 4(6): 700–722.
- Dörr KN and Hollnbuchner K (2017) Ethical challenges of algorithmic journalism. *Digital journalism* 5(4): 404–419.
- Flensburg S and Lomborg S (2021) Datafication research: mapping the field for a future agenda. *New Media & Society*. 14614448211046616.
- Flew T, Spurgeon C, Daniels A, et al. (2012) The promise of computational journalism. *Journalism Practice* 6(2): 157–171.
- Ford H and Hutchinson J (2019) Newsbots that mediate journalist and audience relationships. *Digital Journalism* 7(8): 1013–1031.
- Friedman B and Nissenbaum H (1996) Bias in computer systems. *ACM Transactions on Information Systems* 14(3): 330–347.
- Fürst S (2020) In the service of good journalism and audience interests? How audience metrics affect news quality. *Media and Communication* 8(3): 270–280.
- Gillespie T (2014) The relevance of algorithms In: Gillespie T (ed), *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: The MIT Press, pp. 167–194.
- 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): 50–59.
- Gutierrez Lopez M, Porlezza C, Cooper G, et al. (2023) A question of design: strategies for embedding AI-driven tools into journalistic work routines. *Digital Journalism* 11(3): 484–503.



- Gutierrez-Lopez M, Missaoui S, Makri S, et al. (2019) Journalists as design partners for AI. In: Proceedings of the CHI 2019 ACM Conference on Human Factors in Computing Systems, 04–09 May 2019, Glasgow, UK.
- Gynnild A (2014) Journalism innovation leads to innovation journalism: the impact of computational exploration on changing mindsets. *Journalism* 15(6): 713–730.
- Haapanen L (2022) *Media Councils and Self-Regulation in the Emerging Area of News Automation*. Helsinki: Council for Mass Media in Finland.
- Hannaford L (2015) Computational journalism in the UK newsroom: hybrids or specialists? *Journalism Education* 4(1): 6–21.
- Helberger N (2019) On the democratic role of news recommenders. *Digital Journalism* 7(8): 993–1012.
- Ignatow G and Mihalcea R (2018) *An introduction to text mining*. London: Sage.
- Karlsen J and Stavelin E (2014) Computational journalism in Norwegian newsrooms. *Journalism Practice* 8(1): 34–48.
- Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. London: Sage.
- Klein S (2015) Antebellum data journalism: or, how big data busted Abe Lincoln. Available at <https://www.propublica.org/nerds/antebellum-data-journalism-busted-abe-lincoln>.
- Koenen E, Schwarzenegger C and Kittler J (2021) Data(fication): “Understanding the World through data” as an everlasting revolution. In: Balbi G, Ribeiro N, Schafer V, et al. (eds), *Digital Roots: Historicizing Media and Communication Concepts of the Digital Age*. Berlin, Boston: De Gruyter Oldenbourg, pp. 137–156.
- Komatsu T et al. (2020). *AI should embody our values: Investigating journalistic values to inform AI technology design*. In: *NordiCHI '20: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*. (11) : New York, USA.
- Kotenidis E and Veglis A (2021) Algorithmic journalism—current applications and future perspectives. *Journalism and Media* 2(2): 244–257.
- Lewis SC (2015) Journalism in an era of big data: cases, concepts, and critiques. *Digital Journalism* 3(3): 321–330.
- Lewis SC and Westlund O (2015) Big data and journalism: epistemology, expertise, economics, and ethics. *Digital Journalism* 3(3): 447–466.
- Loosen W (2018) Four forms of datafied journalism. Journalism’s response to the datafication of society. In: *Communicative Figurations Working Paper Series*, (18). Bremen: University of Bremen.
- Lyon D (2019) Surveillance capitalism, surveillance culture and data politics. In: Bigo D, Isin E and Ruppert E (eds), *Data Politics*. London: Routledge, pp. 64–77.
- Mayer-Schönberger V and Cukier K (2013) *Big Data: A Revolution that Will Transform How We Live, Work, and Think*. Boston: Houghton Mifflin Harcourt.
- McNair B (1994) *News and Journalism in the UK*. London: Routledge.
- Meier K, Schützeneder J, García Avilés JA, et al. (2022) Examining the most relevant journalism innovations: a comparative analysis of five European countries from 2010 to 2020. *Journalism and Media* 3(4): 698–714.

- Milosavljević M and Vobič I (2019) Human still in the loop: editors reconsider the ideals of professional journalism through automation. *Digital Journalism* 7(8): 1098–1116.
- Montal T and Reich Z (2017) I, robot. You, journalist. Who is the author? Authorship, bylines and full disclosure in automated journalism. *Digital journalism* 5(7): 829–849.
- Mumford L (1967) *Technics and Human Development: The Myth of the Machine* Vol.I. New York: Harcourt, Brace Jovanovich.
- Mumford L (1970) *The Pentagon of Power: The Myth of the Machine* Vol. II. New York: Harcourt, Brace Jovanovich.
- Napoli PM (2014) Automated media: an institutional theory perspective on algorithmic media production and consumption. *Communication Theory* 24(3): 340–360.
- Örnebring H (2010) Technology and journalism-as-labour: historical perspectives. *Journalism* 11(1): 57–74.
- 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): 853–871.
- Pavlik JV (2013) Innovation and the future of journalism. *Digital Journalism* 1(2): 181–193.
- Pocino PV (2021) *Algorithms in the Newsrooms Challenges and Recommendations for Artificial Intelligence with the Ethical Values of Journalism*. Barcelona: Catalan Press Council.
- Porlezza C (2018) Deconstructing data-driven journalism. Reflexivity between the datafied society and the datafication of news work. *Problemi dell'Informazione* 3: 369–392.
- Porlezza C (2019) Data journalism and the ethics of open source. Transparency and participation as a prerequisite for serving the public good. In: Daly A, Devitt K and Mann M (eds), *Good Data*. Amsterdam: Institute of Network Cultures, pp. 189–201. INC Theory on Demand Series.
- Porlezza C (2020) Ethische Herausforderungen eines automatisierten Journalismus. [Ethical challenges of automated journalism]. In: Köberer N, Prinzing M and Debatin B (eds), *Kommunikations- und Medienethik – reloaded?* Baden-Baden: Nomos, pp. 143–158.
- Porlezza C (2023) Promoting responsible AI: a European perspective on the governance of artificial intelligence in media and journalism. *Communications*. online first DOI: [10.1515/commun-2022-0091](https://doi.org/10.1515/commun-2022-0091).
- Porlezza C and Eberwein T (2022) Uncharted territory. Datafication as a challenge for journalism ethics. In: Karmasin M, Diehl S and Koinig I (eds), *Media and Change Management - Enduring the Challenges of a Constantly Changing Landscape*. Cham: Springer, pp. 343–361.
- Porlezza C and Ferri G (2022) The missing piece: ethics and the ontological boundaries of automated journalism. *#ISOJ* 12(1): 71–98.
- Porter TM (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.
- Prior L (2003) *Using Documents in Social Research*. London: Sage.
- Pybus J and Coté M (2021) Did you give permission? Datafication in the mobile ecosystem. *Information, Communication & Society*: 1–19.
- Rogers S (2011) Data journalism at the Guardian: What is it and how do we do it? The Guardian <https://www.theguardian.com/news/datablog/2011/jul/28/data-journalism>.

- Ryfe D (2012). *Can Journalism Survive. An inside Look at American Newsrooms*. Cambridge: Polity.
- Sadowski J (2019) When data is capital: datafication, accumulation, and extraction. *Big data & society* 6(1) (2053951718820549).
- Schapals AK and Porlezza C (2020) Mastering the robots: assessing the impact of newsroom automation on journalistic role conceptions. *Media & Communication* 8(3): 16–26.
- Schwarz A (2004) *The Numbers Game. Baseball's Lifelong Fascination with Statistics*. New York: Thomas Dunne Books.
- Simon FM (2022) Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy. *Digital Journalism*: 1–23.
- Splendore S (2016) Quantitatively oriented forms of journalism and their epistemology. *Sociology Compass* 10(5): 343–352.
- Stark JA and Diakopoulos N (2016, September). Towards editorial transparency in computational journalism. In: Paper presented at the Computation+ Journalism Symposium (Vol. 5).
- Steenen S (2019) Journalism's epistemic crisis and its solution: disinformation, datafication and source criticism. *Journalism* 20(1): 185–189.
- Tandoc EC Jr and Oh SK (2017) Small departures, big continuities? Norms, values, and routines in The Guardian's big data journalism. *Journalism studies* 18(8): 997–1015.
- Thurman N (2019a) Computational journalism. In: Wahl-Jorgensen K and Hanitzsch T (eds), *The Handbook of Journalism Studies*. 2nd edition. New York: Routledge.
- Thurman N (2019b) Personalization of news. In: Vos T and Hanusch V (eds), *The International Encyclopedia of Journalism Studies*. Hoboken, NJ: Wiley.
- Thurman N, Dörr K and Kunert J (2017) When reporters get hands-on with robo-writing: professionals consider automated journalism's capabilities and consequences. *Digital Journalism* 5(10): 1240–1259.
- Thurman N, Lewis SC and Kunert J (2019) Algorithms, automation, and news. *Digital Journalism* 7(8): 980–992.
- Uricchio W (2017) Data, culture and the ambivalence of algorithms. In: Schäfer MT and Van Es K (eds), *The Datafied Society. Studying Culture through Data*. Amsterdam: Amsterdam University Press, pp. 125–138.
- Usher N (2016) *Interactive Journalism. Hackers, Data and Code*. Urbana: University of Illinois Press.
- Van Dalen A (2012) The algorithms behind the headlines: how machine-written news redefines the core skills of human journalists. *Journalism practice* 6(5-6): 648–658.
- Van Dijck J (2014) Datafication, dataism and dataveillance: big data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.
- Van Dijck J, Nieborg D and Poell T (2019) Reframing platform power. *Internet Policy Review* 8(2): 1–18.
- Weinacht S and Spiller R (2014) Datenjournalismus in Deutschland: eine explorative Untersuchung zu Rollenbildern von Datenjournalisten. *Publizistik* 59(4): 411–433.

Wu S, Tandoc EC Jr and Salmon CT (2019) Journalism reconfigured: assessing human–machine relations and the autonomous power of automation in news production. *Journalism studies* 20(10): 1440–1457.

Zamith R (2019) Algorithms and journalism. In: Örnebring H, Chan YY, Carlson M, et al. (eds), *Oxford Encyclopedia of Journalism Studies*. Oxford: Oxford University Press.

### **Author biography**

Colin Porlezza is a Senior Assistant Professor of Digital Journalism with the Institute of Media and Journalism (IMeG) at the Università della Svizzera italiana (USI). Currently, he is also a Research Fellow with the Tow Center for Digital Journalism at Columbia University, NY, and he also holds an Honorary Senior Research Fellow position with the Department of Journalism at City, University London. His research focuses on automated journalism and AI, in particular on issues related to the governance as well as the responsible use of AI in journalism. In addition, he also researches about the datafication and hybridization of digital journalism, as well as on media accountability and transparency.