This is an Accepted Manuscript of a book chapter to be published in Karin Wahl-Jorgensen and Thomas Hanitzsch (Eds.) (2019) *The Handbook of Journalism Studies, Second Edition*. New York: Routledge.

Computational Journalism

Neil Thurman

This chapter considers *computational journalism* to be the advanced application of computing, algorithms, and automation to the gathering, evaluation, composition, presentation, and distribution of news.

Computational news gathering and evaluation can utilize tools that find and filter newsworthy information from social media platforms and document caches and that provide guidance on the credibility of content and contributors. Such tools include Dataminr, which promises to deliver "the earliest tips for breaking news" and claims to be used in more than 400 newsrooms around the world (Dataminr, n.d.).

Computational news composition and presentation can make use of natural language generation and artificial intelligence to generate written and audio-visual news texts, often from data-feeds. Fanta (2017) found that 9 of the 14 — mainly European — news agencies he surveyed were making use of automated news writing, and two others had projects underway.

Examples of the role computing can take in news distribution include automated news personalization — where stories are chosen and prioritized according to individual users' explicitly registered and / or implicitly determined preferences—and news aggregation sites and apps, like Google News, whose algorithms "determine which stories, images, and videos [to] show, and in what order" (Google, n.d.). According to Thurman (2011), by 2009 the online editions of a sample of large, legacy news providers in the UK and U.S. all carried a considerable variety of tools to tailor stories to their users' interests.

Although some of these practices are not new — automated news personalization dates back to at least the 1980s (Thurman, 2019) — it was only from about 2006 that they started to be discussed under the single, collective term of computational journalism. This chapter provides a summary of, and commentary on, academic studies focused on computational journalism that were published or presented before August 2018. The search term 'computational journalism' was used to query Google Scholar, and the records returned were reviewed. The process of choosing which of the more than 1000 items to include was necessarily subjective. Given the focus of this handbook, technical works from the computer science domain were mostly excluded, or mentioned in passing, in favor of literature from the sociological and behavioral sciences and the humanities.

As will be shown, the focus of computational journalism's literature has broadened over time. An initial emphasis on searching for and analyzing data as part of investigative journalism endeavors has faded as automated news writing, novel forms of interactive news presentation, and personalized news distribution have been addressed. There has also been a growing critical engagement, tempering the early, broadly optimistic analyses with more realistic assessments of computation's effects on the practice of journalism, its content, and reception.

The chapter ends with a discussion of how the literature is evolving, addressing new practices — such as "sensor journalism" and interactive chatbots—and also questioning whether computational journalism's technical essence has been adequately addressed by the sociological contributions to its current corpus.

Emergence

Computational journalism is a relatively new term. It was coined in 2006 by Irfan Essa when he organized the first course on the subject alongside Nick Diakopoulos at Georgia Tech (Georgia Tech, 2013). A blog post by Diakopoulos in January 2007 was entitled "What Is Computational Journalism?" and comprised an early attempt at definition (Diakopoulos, 2007). The term caught on. It started to enter academic parlance. An early mention in academic literature came in the PhD thesis of Adam Perer (2008), where he discussed a computational tool called SocialAction that journalists were starting to value for its facilitation of social network analysis.

SocialAction was a tool developed by and for those outside journalism—in this case by computer scientists for "researchers" (SocialAction, n.d.)—which attracted interest from those within journalism, who used it, for example, to analyze and visualize the social networking links between those implicated in the use and supply of performance-enhancing drugs in baseball (Perer & Wilson, 2007). Collaborations between journalists and technologists followed, and it was one such collaboration that occasioned the use of the term in the pioneering Computational Journalism course taught at Georgia Tech (Perer, 2008, p. 126).

At least one other U.S. university soon followed the Georgia Tech example. In 2009 Duke University appointed Sarah Cohen as Knight Professor of the Practice of Journalism and Public Policy to lead a "computational journalism initiative" ("Washington Post journalists", 2009). At Duke, computational journalism was seen as a way to "help renew watchdog coverage" by "combining traditional public records and database work with new methods and tools from other disciplines" (ibid.). Cohen's background in "computer assisted investigative journalism" was seen as being an "ideal match" for Duke's initiative, which included wanting to develop open-source reporting tools that would "help lower the costs to journalists of discovering and researching stories" (ibid.).

Here was a point, then, at which computer-assisted reporting (CAR) was perceived as having evolved into something else, when developments in journalism's deployment of computers meant that the long-established term CAR no longer seemed adequate and a new term seemed necessary.

The Need for a New Term

What was it that called for a new term? Most writers in this area acknowledge that computers have had a long history in journalism. Anderson and Caswell (2019) describe how CBS News used a computer to predict the outcome of a presidential election in 1952, and what is known as computer-assisted reporting has been around since at least the 1960s, when Philip Meyer was using computers to investigate stories, including the 1967 Detroit riots (Bowen, 1986). In the 1970s Elliot Jaspin was using relational databases for news discovery, a method that allowed him, for example, to discover convicted drug dealers driving school buses. He later founded an organization that became the National Institute for Computer-Assisted Reporting (Cohen, Hamilton, & Turner, 2011).

Various writers have sought to define the distinction between computer-assisted reporting and computational journalism. Hamilton and Turner (2009) said that CAR "tended to be the province of a specialized subset of investigative reporters", while computational journalism tools "will also be adopted by citizen journalists, non-profit news outlets, and NGOs working on government accountability" (p. 16). Flew, Spurgeon, Daniel, and Swift (2012) made the same point. CAR, they wrote, involved "journalism as a practice that could only be undertaken by those officially sanctioned as journalists" (p. 160). Nick Diakopoulos (2011) wrote that computational journalism was inclusive of computer-assisted reporting but was "distinctive in its focus on the processing capabilities" of the computer. Flew et al. cited Miller and Page (2007) in conceiving of computation as a phenomenon that involves "searching, correlating, filtering, identifying patterns, and so on" (p. 158). These activities weren't new, the authors allowed, but could be performed by computational devices "with greater speed and accuracy" (p. 158). Coddington (2015) suggested that "computational journalism goes beyond CAR in its focus on the processing capabilities of computing, particularly aggregating, automating, and abstracting information" (p. 336). He emphasized "the application of computing and computational thinking" to how information is gathered, interpreted, and presented, contrasting this approach with "the journalistic use of data or social science methods more generally" (p. 335). For Skowran (quoted in Claussen, 2009, p. 136), "automation" is a distinguishing characteristic of computational journalism. For Pulimood, Shaw, and Lounsberry (2011), computational journalism is distinguished from CAR by its "more sophisticated approach to applying algorithms and principles from computer sciences and the social sciences to gather, evaluate, organise and present news and information".

Watchdog Journalism

As we have seen, some early conceptions of computational journalism involved journalism's watchdog function, and the first substantive attempt to define the field of computational journalism was a report by James T. Hamilton and Fred Turner (2009) that emerged from a summer workshop organized by the Center for Advanced Study in the Behavioral Sciences at Stanford University and that saw the potential of computation in granting the watchdog keener eyes. In this report, the authors foregrounded the potential they saw in computation to offer reporters "new techniques with which to pursue journalism's long-standing public interest mission" (p. 2). Computational journalism, they wrote, was a new field that could emerge from the convergence of work in computer science, social science, and journalism. They defined it as "the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism" (p. 2).

Watchdog journalism, by their definition, sought to "hold leaders accountable, unmask malfeasance, and make visible critical social trends". It was a means of providing citizens with "the information they need to make many important choices" (p. 2). The authors were idealistic about the role that computational journalism might play in this area. Computational journalism, they said, might create "new blendings of audience, reporter, and commentator … [that might] grow the audience for watchdog journalism and enhance the involvement of citizens in the democratic watchdog process" (p. 9).

Two years later, in conjunction with Sarah Cohen (Cohen et al., 2011), they restated their optimism about the field's accountability potential: about a possible increase in "the public's ability to monitor power" (p. 66). They envisioned it as helping to "level the playing field between powerful interests and the public" (p. 71). Here, then, in these early works on the subject, was an excitement about how computational journalism's news discovery and data questioning potential might make it harder for those in society who were doing harm to hide.

News Discovery

From today's perspective, Hamilton and Turner (2009) set the boundaries of the field relatively narrowly. They envisaged the field as enabling "reporters to explore increasingly large amounts of structured and unstructured information as they search for stories" (p. 2). For these writers, computational journalism built on the tradition of computer-assisted reporting. It was about searching for and analyzing data. They admitted that their take on the field was provisional, and that the field might evolve in unforeseen ways, and they did speculate about the part that computation might play in the later parts of the news cycle, seeing possibilities for a more interactive and personalized news, but their focus was on computational tools being used in the news cycle's early phase, for news discovery rather than for news composition or distribution.

Much of what Cohen et al. (2011) had to say also related to news discovery and the power of computation in searching through data and unearthing newsworthy elements. Flew et al. (2012) saw computation as taking some of the menial toil out of the journalistic role. The utility value of computational journalism, they said, lay in its ability to free "journalists from the low-level work of discovering and obtaining facts", leaving them to focus on "the verification, explanation and communication of news" (p. 167). Here, then, was a journalism that could involve less drudgery and more depth.

Hamilton and Turner (2009) quoted the work of Sarah Cohen in detailing some of the forms that computational news discovery might take. They talked, for example, of computational tools that extract and visualize data from the PDFs that public bodies release as a result of freedom of information requests; from audio or video files; and from local blogs and press releases. They envisioned some degree of automation, with the software able to "scan" and make decisions based on relevance and timing and also provide context with reference to a reporter's previous work.

While the first writers on this subject talked of the potential for computational discovery tools, or of tools developed outside journalism that journalists might be able to find a use for, later writers were able to discuss computational discovery tools developed specifically for journalists. Nick Diakopoulos, Munmun De Choudhury, and Mor Naaman (2012), for example, described the development of SRSR ("Seriously Rapid Source Review"), a system for filtering and assessing the verity of sources found through social media by journalists. Molina (2012) described a system called VSAIH that looked "for news in hydrological data from a national sensor network in Spain" and created "news stories that general users can understand". Hassan et al. (2014) described their FactWatcher system: "It helps journalists identify data-backed, attention-seizing facts which serve as leads to news stories". Schifferes et al. (2014) described a tool — SocialSensor — built for journalists and designed to help "quickly surface trusted and relevant material from social media — with context" (SocialSensor, n.d.). More recently, as Hamilton and Turner envisaged (2009), we have seen computational tools built to help journalists extract data from press releases. For example, "Madi" is a prototype service that automatically scans press releases to provide journalists with background information about the organizations and people mentioned (Zoon, van Dongen, & Lino, 2018).

Widening the Scope

As has been established, many early studies concentrated on the value of computational tools to the process of news discovery, though they did sometimes mention — if only to then dismiss — their application in other areas. Hamilton and Turner (2009) declared that although "the phrase computational journalism carries for some the suggestion of robotic reporters", computational tools were tools "to supplement rather than substitute for efforts by reporters", and their function would be confined to unearthing data and ideas that reporters would then submit to further exploration (p. 12). Later writers and practitioners have extended definitions of computational journalism to include parts of the news cycle beyond news discovery. Diakopoulos (2011), for example, described the potential for computation in news "dissemination and public response", including "personalization and … recommender systems", as well as in the "communication and presentation" of news. The examples he gives in this latter category are to do with interactive data graphics and newsgames, but we should also include machine-generated news content, otherwise known as "automated journalism", which, by 2012, was already being seen as pushing computational journalism into a "new phase" (van Dalen, 2012).

Presentation and Visualization

In describing how computation has been and could be used to change the presentation of news, Diakopoulos (2011) was echoing and anticipating the contributions of other practitioners and theorists. One of the earliest uses of the term *computational journalism* was in Michael Danziger's (2008) Master's thesis where he used it in the context of the production of interactive graphics and data visualizations (p. 71). Some have seen visualization as one of the fundamental characteristics of computational journalism. Karlsen and Stavelin (2014), in seeking to define computational journalism via four factors, talked of a formal factor, which "is most often information visualizations or info graphics" (p. 36). This expanded role for the visual dimension of news has largely been seen as a welcome development. Flew et al. (2012) stated that "data visualizations and graphics can help both readers and journalists cut through dense information in an efficient way" (p. 166). Such visualizations, they said, could be used to help journalists "better understand or refine a story" or for presenting information to readers more powerfully (p. 167). Hamilton and Turner (2009) discussed a visualization tool called "Jigsaw: Visualization for Investigative Analysis", which had been developed for analysts and researchers but which they thought might be of use to journalists. It offered "a visual representation of the connections among individuals and entities that may be mentioned across many different sets of documents" (p. 10). Flew et al. suggested that a potent way of presenting the news may involve granting readers themselves access to data sets and visualization tools: "Such practice would allow readers to humanise or localise what may otherwise be large, incomprehensible sets of data" (p. 167). Something like this eventually came to pass. Wu, Marcus, and Madden (2013) wrote about a tool called MuckRaker, which "provides news consumers with datasets and visualizations that contextualize facts and figures in the articles they read".

New variants of visualization began to emerge. Pavlik and Bridges (2013) considered augmented reality to be "part of a broader emerging field known as computational journalism (CJ)" and discussed how "digital technology might transform the content of journalism through augmented reality". They saw potential for augmented reality in creating media interfaces for those with disabilities, and also hoped that it might make digital journalism more attractive to those news consumers, especially young ones, who had become "disengaged from traditional news media in favor of social media and other newer devices".

Automated Journalism

Although not usually visually distinct from traditional — manually produced — forms of news, so-called automated journalism has become a widely discussed sub-genre of computational journalism. Defined by Carlson (2015, p. 416) as "algorithmic processes that convert data into narrative news texts with limited to no human intervention beyond the initial programming", automated journalism was anticipated as early as 1965 in Michael Frayn's satirical novel, *The Tin Men* (Frayn, 1965). Although it took several decades for Frayn's fantasy to become a reality, automation has, now, been used in the production of written news texts for some time (see, e.g., Dickey, 2014; Dörr, 2015; Gregory, 2017; Lichterman, 2017; Schonfeld, 2010; Young & Hermida, 2015).

The use of automation in the creation of written news texts has been the subject of a number of academic articles. These have examined how the technology has been discussed in the popular press (Carlson, 2015; van Dalen, 2012); how some of the third-party service providers present themselves in public (Carlson, 2015); reactions of journalists who have used the technology first-hand (Thurman, Dörr, & Kunert, 2017; Young & Hermida, 2015); the legal and ethical issues raised (Dörr & Hollnbuchner, 2017; Ombelet, Kuczerawy, & Valcke, 2016); and strategic, business, and labor considerations (Kim & Kim, 2017; Cohen, 2015). There have also been a number of more theoretical contributions — for example on the effect of automation on journalism's ideology (Linden, 2017) and legitimacy (Carlson, 2017) — as well as case studies focusing on the use of automation in news agencies (Fanta, 2017; Marconi & Siegman, 2017) and individual news outlets (Young & Hermida, 2015).

One strand of research has focused on audiences' opinions about written news texts produced with the help of automation, or labeled as such (Clerwall, 2014; Graefe, Haim, Haarmann, & Brosius, 2016; Graefe, Haim, & Diakopoulos, 2017; Haim & Graefe, 2017; Jung, Song, Kim, Im, & Oh, 2017; van der Kaa & Krahmer, 2014; Waddell, 2018; Zheng, Zhong, & Yang, 2018). Most of these studies' findings have shown few, or minor, differences in the way readers perceive human-written and "automated" texts (see, e.g., Clerwall, 2014; van der Kaa & Krahmer, 2014; Haim & Graefe, 2017). However, the research methods used in some of these studies raise questions about the validity of their results, and future studies should ensure that the human and "automated" texts being compared are from the same journalistic genre, cover the same events, and are on topics familiar to respondents and in their native tongue.

Who — or what — is, and should be, credited as the author of automated journalism has been considered by Montal and Reich (2017). They examined how transparent the authorship of the automated journalism published by 12 news organizations was. They found that "most of the studied organizations have some level of transparency: full, partial or low" but that "the identity of the author", where it was not fully human, was inconsistently attributed, ranging from "the software vendor, to the news organization, or the algorithm (bot) itself". They also found "discrepancies between the perceptions of key figures and experts in media organizations pioneering the use of automated journalism and their actual practices concerning bylines and full disclosure".

This, they concluded, "emphasizes the fundamental need for a detailed, comprehensive and transparent bylining and disclosure policy" in the context of automated journalism. In order for this to happen, they suggest that where content is produced without the involvement of a human journalist, the software vendor or the programmer should be attributed. Where content is produced through collaboration between journalists and algorithms, they suggest that the human journalist should be credited but that the objects created by the algorithm should be identified.

News Distribution and Personalization

We have seen, then, how computational tools have been applied to news discovery and to its creation and presentation. Computation has been applied also to news distribution and has allowed the advent of personalized news: news tailored to the preferences of individual users by “explicitly registered” and/or “implicitly determined” means (Thurman, 2011). Nicholas Negroponte’s (1995) “The Daily Me” is often mentioned as an early conception of this phenomenon, though in fact Jules and Michel Verne (1889) had imagined a personalized news service over a hundred years earlier. Some writers have been positive about such an idea. Some less so. As the idea became reality, with personalized news being provided by both traditional media organizations and social media platforms such as Facebook, concerns began to arise about the creation of filter bubbles (Pariser, 2011), with news consumers potentially foregoing exposure to important information and alternative viewpoints, with consequences for the functioning of democracies. While many writers, therefore, had been optimistic about computational journalism’s potential for opening the public’s eyes to the information crucial to democratic health, here was a form of computational journalism accused of doing the opposite. There is disagreement, however, about just how blinkered the populations produced by personalization actually are. Work by Zuiderveen Borgesius et al. (2016), for example, played down such concerns, stating that “personalised content does not constitute a substantial information source for most citizens”. They did allow, though, that “if personalisation technology improves, and personalised news content becomes people’s main information source, problems for our democracy could indeed arise”.

A Sociological Approach

By some accounts, much of the early literature on computational journalism focused on the tools that were being, or could be, built, and the benefits they might bring. Such a focus on what Diakopoulos (2017) has called "tooling" has been subject to criticism for not examining "larger social, political, organizational, and cultural currents in journalism" (Anderson, 2011). Anderson (2011) has advocated for "a more interdisciplinary and externalist perspective on computational journalism research" (p. 5). This, he suggested, could happen, in part, through the application of Schudson's (2005) political, economic, organizational, and cultural approaches.

In starting to flesh out how such applications might develop, he stated, for example, that an economic approach could attempt to "correlate forms of computationally enhanced news production with levels of institutional economic capital" (p. 10), which might show that "certain technologically focused innovations appeared out of reach for less wealthy news organizations" (p. 10).

Anderson also proposed that the "actual role played by materiality and technology in the processes of journalism" should be accounted for (p. 15). In this last suggestion, Anderson has allies. For example, Primo and Zago (2015) have argued that, in journalism studies, technology is often "portrayed as an external force (influence) that impacts humans and what humans produce", and suggested that such binary strategies "artificially fragment journalism, reducing what is an entangled network to opposing poles". They suggested that "technological artifacts and other objects also do journalism. Thus, besides 'who,' we also need to ask 'what' does journalism". Lewis and Westlund (2015) concur, saying that "during the past two decades, journalism studies scholars have paid special attention to the role of technology in news work" but that "this line of research has given greater emphasis to human-centric considerations" and not sufficiently acknowledged "the distinct role of technology and the inherent tension between human and machine approaches" (p. 20). They say there is "an opportunity for developing a sociotechnical emphasis in journalism studies" (p. 21) which would acknowledge "the extent to which contemporary journalism is becoming interconnected with technological tools, processes, and ways of thinking as the new organizing logics of media work" (p. 21).

The Economic Lens

In calling for a more critical approach to computational journalism, Anderson (2011) suggested that an economic lens is one that may be beneficial, and highlighted what he believed to be an absence of "work done on the relationship between economic resources and computational journalism", for example how "different institutionally specific resources constrain the options available to various news outlets and industry segments".

The literature on computational journalism, even early on, considered economic factors, although, it is true, sometimes putting a sharper, or even exclusive, focus, on the cost benefits it might bring rather than the inequalities it might promote. Flew et al. (2012), for example, in talking of how investigative journalism could involve the laborious checking of thousands of documents, and how computation could spare journalists such lengthy toil, talked of savings in time and savings in cost. Cohen et al. (2011) considered computational journalism against the backdrop of increasing financial difficulty faced by traditional news providers: the pressures placed on public-affairs reporting by "the decline in revenue and reporting staff in traditional news organizations". This, they stated, was "where the field of computational journalism can help the most" (p. 68).

However, scholars, even early on, were aware of financial complications in this idealistic picture, with, for example, Hamilton and Turner (2009) stating that tools would need "a very low cost of acquisition, since local papers and online news providers will be hard-pressed to make investments in accountability coverage" (p. 12). Flew et al. (2012) talked of "significant software and technology start-up costs" involved in the adoption of computational journalism in news organizations (p. 165). Diakopoulos (2017) wrote that the lower costs associated with computational journalism "do not always materialize", and Sylvian Parasie's (2015) case study of a journalism project that developed algorithms and databases in the service of its investigation into seismic safety standards in California showed how the time and costs involved could be problematic for other news suppliers in the current financial climate for journalism.

Various solutions have been proposed to make computational journalism tools more affordable to journalists and publishers, including alternative funding methods, open-source software, crowd-sourcing, and entrepreneurial initiatives on the part of newsrooms. Hamilton and Turner (2009), and Cohen et al. (2011), spoke of the need for funding to come from outside journalism, with media organizations reluctant to invest in areas that are not "readily monetized" (Hamilton & Turner, 2009, p. 13). One outside source highlighted by Flew et al. (2012) was the Sunlight Foundation, a non-profit, nonpartisan organization with a goal of scrutinizing government

> that has arisen in the light of the plethora of US data made publicly available under initiatives of greater government openness and transparency … [and] has been involved in both the creation of freely available tools and websites that enable individuals and communities to access and engage with government information (p. 165).

We are also starting to see some entrepreneurial activities, with news organizations developing computational tools themselves. Reuters has built, in house, a tool called "Tracer," which enables "journalists to spot and validate real news in real time on Twitter" (Reuters, n.d.), and the *Washington Post* is behind a suite of publishing tools, including Clavis, "a personalization engine powered by natural language processing" (Arc Publishing, n.d.), which it sells to other publishers. Although such developments are in line with Diakopoulos's (2017) call for the journalism industry to develop its own tools, the exclusive access Reuters has to its Tracer product, and the cost of using the *Washington Post*'s suite of tools — between $10,000 and $150,000 a month (Ingram, 2017) — are not quite in the spirit of Diakopoulos's call for news organizations to be "cultivating communities around … open source tools".

More in the spirit of Diakopoulos's call was *The Guardian*'s use of crowd-sourcing to search through a huge number of documents relating to MPs' expenses, which, according to Flew at al. (2012, p. 163), was achieved at a low cost. Andersen (2009) says that the necessary software took a developer one week to build and that an additional £50 was required to "rent temporary servers".

Algorithmic Transparency and Accountability

Calls — for example by Anderson (2011), Lewis and Westlund (2015), and Primo and Zago (2015) — for closer attention to be paid to the distinct role played by technological artefacts in computational journalism are also starting, slowly, to be addressed, as attempts are made to make the inner workings of algorithms more transparent. Nicholas Diakopoulos and Michael Koliska (2017) provide some examples of where this has happened, for example NYTimes.com blogging about how its personalized news recommendation engine works and the open-sourcing of data and code used to build some of the data-driven articles (p. 810) published by BuzzFeed. Thurman et al. (2016) have shown how one tool, built to help journalists identify trending news stories in social media, relies mostly on metropolitan men in the mainstream media as inputs and prioritizes stories about people, places, and organizations that have been subject to short-term spikes of interest on social media. Although such characteristics are open to criticism, Thurman et al. (2016) emphasize how algorithms often mirror established practices and stress the importance of changes outside code, for example to the "demography of the journalism profession". Other research has described attempts to build transparent news filtering/recommender algorithms that focus on journalistic value (Song, Oh, & Jung, 2018).

Such examples are, however, relatively few and far between. Part of the reason, suggest Diakopoulos and Koliska (2017), is "a lack of business incentives for disclosure" and "the concern of overwhelming end-users with too much information". This latter concern may have some empirical basis. In a pilot study, Graefe et al. (2017) found that increased transparency about the authorship of an "automated" news story was correlated with lower levels of audience appreciation for the story's credibility. In spite of such possible obstacles to transparency, Diakopoulos and Koliska (2017) have outlined a transparency framework for computational journalism algorithms that covers the data they use, how the data are modeled and inferences made, as well as how "any transparency information revealed about an algorithm" could "ultimately take some 'tangible or visual' form in order to be presented to the end-user".

While Diakopoulos and Koliska's (2017) transparency framework was developed with the algorithms used in computational journalism in mind, it could equally apply to algorithms used in any context. Indeed, Diakopoulos (2015) has suggested this should happen and, in doing so, proposed extending the scope of computational journalism to include the journalistic investigation of algorithms, foregrounding the "journalism" in "computational journalism" "by making computation its object". This "algorithmic accountability reporting" would, he suggested, seek to "articulate the … biases, and influences" embedded in computational artifacts that play a role in society. Diakopoulos proposed that algorithmic power could be analyzed by looking at the decisions algorithms make, including how they prioritize, classify, associate, and filter information. In order to facilitate such analysis, the creators of algorithms could disclose information about how they work, although he acknowledged that the business and security interests of commercial and governmental organizations might prevent this from happening. When this is the case, Diakopoulos suggested that a “different, more adversarial approach” could be employed, involving “reverse engineering”. He provided an analysis of the “opportunities and limitations of a reverse engineering approach to investigating algorithms” through interviews with journalists who had done just that, concluding that reverse engineering can “elucidate significant aspects of algorithms such as censorship.”

Conclusion

The practice of computational journalism — the advanced application of computing, algorithms, and automation to the gathering, evaluation, composition, presentation, and distribution of news—is not new. Since as far back as the 1960s, reporters have been employing computers to interpret information as part of their investigative journalism. The use of computers’ processing capabilities to automate the presentation of news goes back decades too, with news personalization deployed by commercial providers since at least the 1980s (Thurman, 2019). Between the interpretation of information and its presentation as news, there is, of course, a compositional process, where news items are written and edited. Although some of the early literature on computational journalism (Hamilton & Turner, 2009) played down the potential of computing in this phase of the news production cycle, so-called automated journalism is now firmly established. Computational news gathering—at least at scale—also took a while to take off, but has now done so, driven by the increasing volumes of digital data, including on social media platforms, that contain potentially newsworthy nuggets.

Although such practices have been growing in prevalence for decades, it was not until the mid-noughties that they began to be discussed under a single, collective term. The focus of such discussions in the early computational journalism literature was on the use of computing to explore and interpret data, with a strong stress given to journalism's "watchdog," "accountability," and "monitorial" functions. The computer-assisted reporting backgrounds of some of those early writers, and their location within normatively orientated U.S. journalism schools, offers some explanation, perhaps, for this early emphasis. There was also, initially, optimism about the potential for computational journalism, perhaps attributable to the attention the literature paid to making and doing.

As the field has developed, its literature has more fully reflected the variety of computational journalism practices and become more realistic about its potential to, for example, "level the playing field between powerful interests and the public" (Cohen et al., 2011, p. 71). The information exploration and interpretation applications emphasized in the literature early on remain an important avenue for research and practice. Work on—and about—tools to help journalists explore, extract, and visualize information continues, but there has been a growing emphasis on verification (see, e.g., Fletcher, Schifferes, & Thurman, 2017), a result of the increasing volume of misleading and manipulated information in circulation, both from social media users and official sources.

Alongside its ongoing interest in information discovery, the computational journalism literature has expanded to reflect the increased use of computation and automation in the composition of news. This strand of research focused, initially, on traditionally formatted, static, written news texts, but is now starting to encompass automated, interactive news chatbots (see, e.g., Jones & Jones, 2018b and Ford & Hutchinson, 2018) and the automation of short-form news video (Thurman, Schulte-Uentrop, Rogge, & Krueger, 2018). It is also starting to reflect the use of automated journalism among news organizations at the local level (see, e.g., Alabaster, Silcock, & Chadha, 2018) and the use of sensors embedded in the real world as a source of data driving the composition and distribution of automated news items. Examples of this "sensor-journalism" or "sensor-telling" have covered topics such as pollution and animal welfare (Vicari & Weiss, 2018).

The use of computation in personalized news distribution—and the academic and popular discourse around it — has a substantially longer history than sensor journalism. Whereas some of the pioneering authors on the topic took a normative approach, contemporary writings are more evidential: exploring whether and why news consumers think automated personalization is a better way to get news than selection by journalists and editors (Thurman, Moeller, Helberger, & Trilling, 2018); questioning received wisdom on the existence of filter bubbles (see, e.g., Zuiderveen Borgesius et al., 2016); and even asking whether recommendation engines might promote, rather than limit, diverse news exposure (Helberger, Karppinen, & D'Acunto, 2018).

At the same time as embracing a wider range of practices, the computational journalism literature has also become more sophisticated in its methods and more realistic — critical even — about computation's effects on the practice of journalism, its content, and reception. Hopes that computational news discovery would make it harder for those in society who are doing harm to hide have been tempered by a realization that the very tools being built to enable such discovery may be unaffordable to some publishers, surveil citizens in a "stalker-esque" fashion (Thurman, 2018), push a popularist news agenda (ibid.), and have — or at least reflect existing — biases in their sourcing practices and determinations of newsworthiness (Thurman et al., 2016).

There are criticisms too of computational news composition, including about the one-dimensional nature of the quantitative feeds that much of it relies on (Thurman et al., 2017), the dumbed-down nature of some of its output (Ford & Hutchinson, 2018), the effects of the almost unlimited volumes of news and information it can propel into the public sphere, and the consequences — economic and ethical — of journalistic expertise being embodied in software platforms that are available to anybody, whatever their motivation or institutional affiliation — or lack of.

Such criticisms are, however, often constructively made, accompanied by concrete suggestions about how, for example, to make computational journalism's algorithms more transparent and accountable. The emerging computational journalism literature also reminds us that the consequences of computation for journalism may be less dramatic, and unfold more slowly, than some have predicted (see, e.g., Linden, Sirén-Heikel, Haapanen, & Moring, 2018; Schapals, 2018; Ferrer-Conill & Clerwall, 2018; Milosavljević & Vobič, 2018; and Stray, 2018) and that it is merely the most recent manifestation of a longer history of "quantitative journalism" (Anderson, 2015).

Computational journalism was a latecomer to the journalism studies table, relatively inscrutable, even to itself. Developing initially with a relatively narrow and somewhat practice-orientated bent, it has begun to mature, recognizing the full spectrum of its interests and some of its own limitations. Whether, in the future, it will hold together or fragment remains to be seen. Some believe the literature to be "utilitarian, analytical, and theoretical … primarily sociological rather than technical" and have called for a complementary approach based on first principles (Anderson & Caswell, 2019). If such an approach takes hold, we can expect to see more literature in the mold of Jones and Jones's (2018a) and Caswell's (2018) that, as Anderson and Caswell (2019) suggest, participates in computational journalism on its own terms and advances it as a technological practice.

References

Alabaster, J., Silcock, W. B., & Chadha, M. (2018, May). *Local at scale: Examining the automation of hyperlocal news*. Paper presented at Algorithms, Automation, and News conference, LMU Munich, Germany.

Andersen, M. (2009, June 23). Four crowdsourcing lessons from the *Guardian*'s (spectacular) expenses-scandal experiment. *NiemanLab*. Retrieved from http://www.niemanlab.org/2009/06/four-crowdsourcing-lessons-from-the-guardians-spectacular-expenses-scandal-experiment/

Anderson, C. W. (2011). Notes towards an analysis of computational journalism. *HIIG Discussion Paper Series No. 2012–1*. Retrieved from http://dx.doi.org/10.2139/ssrn.2009292

Anderson, C. W. (2015). Between the unique and the pattern: Historical tensions in our understanding of quantitative journalism. *Digital Journalism*, *3*(3), 349–363. doi:10.1080/21670811.2014.976407

Anderson, C. W., & Caswell, D. (2019). Computational journalism. In T. Vos & F. Hanusch (Eds.), *The international encyclopedia of journalism studies*. Hoboken, NJ: John Wiley & Sons.

Arc Publishing (n.d.). Arc products. Retrieved from https://www.arcpublishing.com/products/

Bowen, E. (1986, July 7). Press: New paths to buried treasure: Computers are revolutionizing investigative journalism. *Time* magazine.

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. (2017). Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism. *New Media & Society*. Advance online publication. doi:10.1177/1461444817706684

Caswell, D. (2018, May). *Structured journalism and the semantic units of news*. Paper presented at Algorithms, Automation, and News conference, LMU Munich, Germany.

Claussen, D. S. (2009). How one would really “blow up” a j-school curriculum. *Journalism & Mass Communication Educator, 64*(2), 133–136.

Clerwall, C. (2014). Enter the robot journalist: Users' perceptions of automated content. *Journalism Practice*, *8*(5), 519–531.

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. doi:10.1080/21670811.2014.976400

Cohen, N. S. (2015). From pink slips to pink slime: Transforming media labor in a digital age. *The Communication Review*, *18*(2), 98–122.

Cohen, S., Hamilton, J. T., & Turner, F. (2011). Computational journalism. *Communications of the ACM*, *54*(10), 66–71.

Danziger, M. (2008). *Information visualization for the people* (Master’s thesis, Massachusetts Institute of Technology). Retrieved from http://cmsw.mit.edu/wp/wp-content/uploads/2008/05/146381107-Michael-Danziger-Information-Visualization-for-the-People.pdf

Dataminr (n.d.). News. Retrieved from https://www.dataminr.com/news

Diakopoulos, N. (2007, January 13). What is computational journalism? [Blog post] Retrieved from http://gt-cj.blogspot.com/

Diakopoulos, N. (2011). A functional roadmap for innovation in computational journalism. Retrieved from http://www.nickdiakopoulos.com/wp-content/uploads/2007/05/CJ_Whitepaper_Diakopoulos.pdf

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, *3*(3), 398–415. doi:10.1080/21670811.2014.976411

Diakopoulos, N. (2017). Computational journalism and the emergence of news platforms. In B. Franklin & S. Eldridge II (Eds.), *The Routledge companion to digital journalism studies* (pp. 176–184). Abingdon: Routledge.

Diakopoulos, N., De Choudhury, M., & Naaman, M. (2012). Finding and assessing social media information sources in the context of journalism. *CHI'12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, Texas, USA*, 2451–2460.

Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, *5*(7), 809–828. doi:10.1080/21670811.2016.1208053

Dickey, M. R. (2014, March 17). How an *LA Times* reporter got an algorithm to write articles for him. *Business Insider*. Retrieved from http://www.businessinsider.com/quakebot-robot-la-times-2014-3?IR=T

Dörr, K. (2015). Mapping the field of algorithmic journalism. *Digital Journalism*, *4*(6), 700–722. doi:10.1080/21670811.2015.1096748

Dörr, K., & Hollnbuchner, K. (2017). Ethical challenges of algorithmic journalism. *Digital Journalism*, *5*(4), 404–419.

Fanta, A. (2017). Putting Europe's robots on the map: Automated journalism in news agencies. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-09/Fanta%2C%20Putting%20Europe%E2%80%99s%20Robots%20on%20the%20Map.pdf

Ferrer-Conill, R., & Clerwall, C. (2018, May). *The journalistic conveyor belt: Intensities of automation and algorithmic micro-processes in Swedish newsrooms*. Paper presented at Algorithms, Automation, and News conference, LMU Munich, Germany.

Fletcher, R., Schifferes, S., & Thurman, N. (2017). Building the “Truthmeter”: Training algorithms to help journalists assess the credibility of social media sources. *Convergence: The International Journal of Research into New Media Technologies*, 1–16. doi:10.1177/1354856517714955

Flew, T., Spurgeon, C., Daniel, A., & Swift, A. (2012). The promise of computational journalism. *Journalism Practice*, *6*(2), 157–171.

Ford, H., & Hutchinson, J. (2018, May). *Bots and public service media: Democratic bedfellows or an unholy alliance?* Paper presented at Algorithms, Automation, and News conference, LMU Munich, Germany.

Frayn, M. (1965). *The tin men*. London: William Collins Sons & Company Ltd.

Georgia Tech (2013). GVU brown bag seminar: Computational journalism. Retrieved from http://www.gatech.edu/hg/item/182791

Google News (n.d.). How Google News stories are selected. Retrieved from https://support.google.com/googlenews/answer/9005749?hl=en&ref_topic=7688382#

Graefe, A., Haim, M., & Diakopoulos, N. (2017, May). *Should news outlets let their readers know that they are reading automated content? Effects of algorithmic transparency on perceptions of automated news*. Poster presented at the 67th annual meeting of the International Communication Association, San Diego, CA, USA.

Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. (2016). Readers’ perception of computer-generated news: Credibility, expertise, and readability. *Journalism*. Advance online publication. doi:10.1177/1464884916641269

Gregory, J. (2017, July 6). Press Association wins Google grant to run news service written by computers. *The Guardian*. Retrieved from https://www.theguardian.com

Haim, M., & Graefe, A. (2017). Automated news: Better than expected? *Digital Journalism*, *5*(8), 1044–1059.

Hamilton, J. T., & Turner, F. (2009). Accountability through algorithm: Developing the field of computational journalism. Retrieved from http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf

Hassan, N., Sultana, A., Wu, Y., Zhang, G., Li, C., Yang, J., & Yu, C. (2014). Data in, fact out: Automated monitoring of facts by FactWatcher. *Proceedings of the VLDB Endowment, 7*(13), 1557–1560. doi:10.14778/2733004.2733029

Helberger, N., Karppinen, K., & D'Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society, 21*(2), 191–207. doi:10.1080/1369118X.2016.1271900

Ingram, M. (2017). How the *Washington Post* makes money from its competitors. *Fortune*. Retrieved from http://fortune.com/2017/03/13/washington-post-arc/

Jones, R., & Jones, B. (2018a, May). *Atomised news: Experimenting with structure in public service journalism*. Paper presented at Algorithms, Automation, and News conference, LMU Munich, Germany.

Jones, R., & Jones, B. (2018b, May). *Public service chatbots: Automating conversation with BBC News*. Paper presented at Algorithms, Automation, and News conference, LMU Munich, Germany.

Jung, J., Song, H., Kim, Y., Im, H., & 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*(C), 291–298.

Karlsen, J., & Stavelin, E. (2014). Computational journalism in Norwegian newsrooms. *Journalism Practice*, *8*(1), 34–48.

Kim, D., & Kim, S. (2017). Newspaper companies' determinants in adopting robot journalism. *Technological Forecasting and Social Change*, *117*(C), 184–195.

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

Lichterman, J. (2017). Want to bring automation to your newsroom? A new AP report details best practices. *NiemanLab*. Retrieved from http://www.niemanlab.org/2017/04/want-to-bring-automation-to-your-newsroom-a-new-ap-report-details-best-practices/

Linden, C.-G. (2017). Decades of automation in the newsroom: Why are there still so many jobs in journalism? *Digital Journalism*, *5*(2), 123–140.

Linden, C.-G., Sirén-Heikel, S., Haapanen, L., & Moring, T. (2018, May). *So much data, so few stories: Journalistic usability of open data*. Paper presented at Algorithms, Automation, and News conference, LMU Munich, Germany.

Marconi, F., & Siegman, A. (2017). The future of augmented journalism: A guide for newsrooms in the age of smart machines. *Associated Press*. Retrieved from https://insights.ap.org/uploads/images/ap_insights_the_future_of_augmented_journalism.pdf

Miller, J. H., & Page, S. E. (2007). *Complex adaptive systems: An introduction to computational models of social life*. Princeton, NJ: Princeton University Press.

Milosavljević, M., & Vobič, I. (2018, May). *Human-still-in-the-loop: Journalism between automation and professionalization*. Paper presented at Algorithms, Automation, and News conference, LMU Munich, Germany.

Molina, M. (2012). Simulating data journalism to communicate hydrological information from sensor networks. In J. Pavón, N. D. Duque-Méndez, & R. Fuentes-Fernández

(Eds.), *Advances in Artificial Intelligence – IBERAMIA 2012* (pp. 722–731). Heidelberg: Springer.

Montal, T., & Reich, Z. (2017). I, robot. You, journalist. Who is the author? *Digital Journalism*, *5*(7), 829–849. doi:10.1080/21670811.2016.1209083

Negroponte, N. (1995). *Being digital*. London: Hodder & Stoughton.

Ombelet, P.-J., Kuczerawy, A., & Valcke, P. (2016). Employing robot journalists: Legal implications, considerations and recommendations. *Proceedings of the 25th International Conference Companion on World Wide Web*. doi:10.1145/2872518.2890093

Parasie, S. (2015). Data-driven revelation? Epistemological tensions in investigative journalism in the age of “big data”. *Digital Journalism*, *3*(3), 364–380. doi:10.1080/21670811.2014.976408

Pariser, E. (2011). *The filter bubble*. London: Penguin Books.

Pavlik, J. V., & Bridges, F. (2013). The emergence of augmented reality (AR) as a storytelling medium in journalism. *Journalism & Communication Monographs*, *15*(1), 4–59. doi:10.1177/1522637912470819

Perer, A. (2008). *Integrating statistics and visualization to improve exploratory social network analysis* (Doctoral dissertation, University of Maryland, College Park). Retrieved from http://search.proquest.com/docview/304565254?pq-origsite=gscholar

Perer, A., & Wilson, C. (2007, December 21). The steroids social network. *Slate.com*. Retrieved from http://www.slate.com/articles/sports/sports_nut/2007/12/the_steroids_social_network.html

Primo, A., & Zago, G. (2015). Who and what do journalism? *Digital Journalism*, *3*(1), 38–52. doi:10.1080/21670811.2014.927987

Pulimood, S. M., Shaw, D., & Lounsberry, E. (2011). Gumshoe: a model for undergraduate computational journalism education. *Proceedings of the 42nd ACM technical symposium on Computer science education (SIGCSE '11)*, 529–534. doi:10.1145/1953163.1953314

Reuters (n.d.). Reuters news tracer. Retrieved from https://agency.reuters.com/en/insights/articles/articles-archive/reuters-news-tracer-filtering-through-the-noise-of-social-media.html

Schapals, A. K. (2018, May). *The challenges and opportunities of automated journalism: Illuminating the status quo in German newsrooms*. Paper presented at Algorithms, Automation, and News conference, LMU Munich, Germany.

Schifferes, S., Newman, N., Thurman, N., Corney, D., Göker, A., & Martin, C. (2014). Identifying and verifying news through social media: Developing a user-centred tool for professional journalists. *Digital Journalism*, *2*(3), 406–418.

Schonfeld, E. (2010, November 12). Automated news comes to sports coverage via StatSheet. *Techcrunch*. Retrieved from https://techcrunch.com/2010/11/12/automated-news-sports-statsheet/

Schudson, M. (2005). Four approaches to the sociology of news. In J. Curran & M. Gurevitch (Eds.), *Mass Media and Society* (4th ed.) (pp. 172–197). London: Hodder Arnold.

SocialAction (n.d.). SocialAction: Integrating statistics and visualization for social network analysis. Retrieved from http://www.cs.umd.edu/hcil/socialaction/

SocialSensor. (n.d.). Introducing SocialSensor. Retrieved from http://www.socialsensor.eu/

Song, H., Oh, S., & Jung, J. (2018, May). *Toward transparent news algorithms focusing on the journalistic values: A case study of Korea Press Foundation's News Trust Project in South Korea*. Paper presented at Algorithms, Automation, and News Conference, LMU Munich, Germany.

Stray, J. (2018, May). *The state of the art of investigative AI*. Paper presented at Algorithms, Automation, and News conference, LMU Munich, Germany.

Thurman, N. (2011). Making "The Daily Me": Technology, economics and habit in the mainstream assimilation of personalized news. *Journalism: Theory, Practice & Criticism*, *12*(4), 395–415. doi:10.1177/1464884910388228

Thurman, N. (2018). Social media, surveillance, and news work: On the apps promising journalists a "crystal ball". *Digital Journalism*, *6*(1), 76–97.

Thurman, N. (2019). Personalization of news. In T. Vos & F. Hanusch (Eds.), *The international encyclopedia of journalism studies*. Hoboken, NJ: John Wiley & Sons.

Thurman, N., Dörr, K., & 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., Moeller, J., Helberger, N., & Trilling, D. (2018). My friends, editors, algorithms, and I: Examining audience attitudes to news selection. *Digital Journalism*. doi:10.1080/21670811.2018.1493936

Thurman, N., Schifferes, S., Fletcher, R., Hunt, S., Schapals, A. K., & Newman, N. (2016). Giving computers a nose for news: Exploring the limits of story detection and verification. *Digital Journalism*, *4*(7), 838–848.

Thurman, N., Schulte-Uentrop, I., Rogge, C., & Krueger, K. (2018, May). *Automating short-form news videos: Perception studies with US consumers*. Paper presented at Algorithms, Automation, and News conference, LMU Munich, Germany.

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 der Kaa, H., & Krahmer, E. (2014). Journalist versus news consumer: The perceived credibility of machine written news. *Proceedings of the Computation+Journalism Conference, New York*. Retrieved from https://pure.uvt.nl/portal/files/4314960/c

Verne, J., & Verne, M. (1889, February). In the year 2889. *The Forum*. Retrieved from https://www.gutenberg.org/files/19362/19362-h/19362- h.htm

Vicari, J., & Weiss, B. (2018, May). *Sensor driven journalism: Combining reporting with the internet of things*. Paper presented at Algorithms, Automation, and News conference, LMU Munich, Germany.

Waddell, T. F. (2018). A robot wrote this? How perceived machine authorship affects news credibility. *Digital Journalism*, *6*(2), 236–255.

Washington Post journalists to join Sanford faculty. (2009, Summer). *Focus*. Retrieved from http://news.sanford.duke.edu/sites/news.sanford.duke.edu/files/newsletters/focus06_09.pdf

Wu, E. B., Marcus, A., & Madden, S. (2013). Data in context: Aiding news consumers while taming dataspaces. *DBCrowd 2013: First VLDB Workshop on Databases and Crowdsourcing*, 47–50. Retrieved from http://dbweb.enst.fr/events/dbcrowd2013/proceedings.pdf

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), 381–397. doi:10.1080/21670811.2014.976409

Zheng, Y., Zhong, B., & Yang, F. (2018). When algorithms meet journalism: The user perception to automated news in a cross-cultural context. *Computers in Human Behavior*, *86*, 266–275.

Zoon, H., van Dongen, J., & Lino, J. A. (2018, May). *Designing for automated journalism in the Netherlands: First steps and no way back*. Paper presented at the Algorithms, Automation, and News Conference, LMU Munich, Germany.

Zuiderveen Borgesius, F. J., Trilling, D., Möller, J., Bodó, B., de Vreese, C. H., & Helberger, N. (2016). Should we worry about filter bubbles? *Internet Policy Review*, *5*(1). doi:10.14763/2016.1.401