

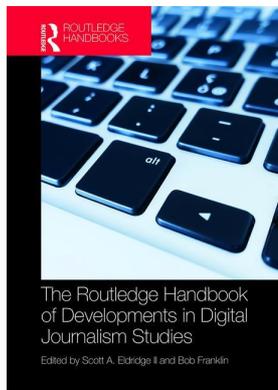
This article was downloaded by: 10.3.98.93

On: 17 Jan 2019

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



The Routledge Handbook of Developments in Digital Journalism Studies

Scott A. Eldridge, Bob Franklin

Algorithms are a Reporter's New Best Friend

Publication details

<https://www.routledgehandbooks.com/doi/10.4324/9781315270449-19>

Carl-Gustav Linden

Published online on: 30 Aug 2018

How to cite :- Carl-Gustav Linden. 30 Aug 2018, *Algorithms are a Reporter's New Best Friend from: The Routledge Handbook of Developments in Digital Journalism Studies* Routledge
Accessed on: 17 Jan 2019

<https://www.routledgehandbooks.com/doi/10.4324/9781315270449-19>

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: <https://www.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

ALGORITHMS ARE A REPORTER'S NEW BEST FRIEND

News automation and the case for augmented journalism

Carl-Gustav Linden

Automated routine news is gradually complementing manually produced journalistic texts, creating emotional stress in many newsrooms. This chapter provides an alternative consideration to the conventional view that automation amounts to an existential threat to the work of journalists. Focusing on opportunities, it discusses how journalists and editors could embrace new media technology built on algorithms in ways that create value and meaning. Theoretically the study is inspired by a social constructionist perspective, arguing that technical, economic, and social factors determine the pace and extent of automation. This chapter turns the usual discussion about job losses on its head with a claim that “algorithms are a journalist’s new best friend”. The real risk is not too much automation but that journalists and the news media falls behind other industries in developing new cognitive technologies because of, for instance, lack of skills, strategy, and scarce financial resources.

In recent years, we have seen new job titles surfacing in newsrooms; suddenly there are news robot trainers, API (Application Programming Interface) editors or algorithm editors working side by side with more traditional journalists (see also Montal and Reich, this volume, Chapter 4). There has been a revolution in the production of news by smart machines thanks to the advancement of Natural Language Generation (NLG) technologies, thus extending human capabilities and creating augmented journalism (Leppanen et al., 2017; Marconi and Siegma, 2017; Scherer, 2010). Journalists are facing the automation of news with mixed feelings, also as a threat to the profession (Linden, 2017; Carlson, 2015; Van Dalen, 2012). “Should We Be Afraid or Excited About Robot Journalism?” asks the Huffington Post (Taibi, 2014), while the advice from *Editor & Publisher* is “Learn to Stop Worrying and Love Robot Journalists” (Tornoe, 2014). The economist David Autor, an expert on the history of automation, criticizes journalists and analysts who “overstate the extent of machine substitution for human labor and ignore the strong complementarities that increase productivity, raise earnings, and augment demand for skilled labor” (Autor, 2014: 130). The computer scientist Ben Shneiderman also believes the discussion around automation has gone off track: “Robots and AI [artificial intelligence] make compelling stories for journalists, but they are a false vision of the major economic changes”, he notes in a survey on AI, robotics, and the future of jobs (Smith and Anderson, 2014: 6). In reality, there has been more talk about news automation than there are actual practical examples (Dörr, 2015) and

one is right to suspect that the debate serves as PR for service providers looking for better-paying customers in industries outside the financially restrained media sector.

Journalists have for decades, mostly unreflectively, been assisted by less controversial algorithms in the newsroom for editing, publishing, and distributing content, especially in broadcast production (Kolkey, 1999; Vigneaux, 1996). The term algorithm refers to a self-contained step-by-step set of operations to be performed, such as calculation, data processing, and automated reasoning – a set of rules that defines a sequence of precise instructions that can be understood by a computer. Adoption of new technology is a key feature of journalism and media work (Pavlik, 2000). However, organizational structure and occupational practice shape the ways in which new technology is adopted and with what effects (Boczkowski, 2005; Dickinson et al., 2013; Young and Hermida, 2014). The discussion among journalists about “robot journalism” is here framed as a recent example of “automation anxiety” (Akst, 2013), which has been a recurrent topic in debates on work and technological development for centuries. Reading older research from the 1980s on computer anxiety among journalists (Shipman, 1986), one is tempted to ask: what is new? The fear of technology seems to be a constant factor in the newsroom (Boczkowski, 2005). Therefore, one can also inquire, if algorithms have been aiding journalists ever since the 1970s, would it be appropriate to take a more nuanced view? The risk of “false alarmism” when discussing technological disruption and labor is imminent (Atkinson and Wu, 2017).

We could start by looking at the jobs issue. Surveys in the United States (Weaver, 2015) show that the number of full-time journalists working for mainstream general-interest news media have declined substantially from about 122,000 in 1992 to about 83,000 in 2013. In contrast, according to the Bureau of Labor Statistics, in 2014, 54,400 people were working as “Reporters, Correspondents, and Broadcast News Analysts” (Bureau of Labor Statistics, 2016), with a forecast that 9%, or 4,400 jobs, will disappear by 2024. However, that comparison does not reflect the changing nature of the occupation, since the full-time qualifier does not quantify the transition from salary to contract work or the increasing share of freelancers and other contributors.

In the United Kingdom, the situation is different: 2016 was a record year, with 84,000 people employed as “journalists, newspaper and periodical editors”, according to official estimates (Office for National Statistics, 2016). This is a dramatic increase from the previous record year, 2013, with 70,000 people. Before the financial crisis struck in 2007 there were 57,000 people officially working as journalists. The proportion of journalists working online is on the rise, with the share of newspaper journalists declining. An estimated 30,000 journalists work wholly or partly online; however, many bloggers are excluded from this count, “along with others whose journalistic identity is complex” (Hargreaves, 2016: 4).

The history of automated storytelling is more than 40 years old (Meehan, 1977) and since the 1960s automatic text summaries have been applied, for example, to weather forecasts (Glahn, 1970), and in the late 1990s to sports, medical, and financial reports. Despite these opportunities, technologists predict that creative occupations such as journalism are the least susceptible occupations to workplace automation (Bakhshi et al., 2015). It seems that where journalism jobs have been disappearing, the prime reasons were changing consumer behavior and media business models, not automation.

Previous research

In recent years software-generated news has gradually become part, albeit a small part, of the news-making ecosystem, thus blurring the boundaries between journalism, computer science and statistics. Popularly labeled “robot journalism” (Clerwall, 2014; Latar, 2014; Levy, 2012; Matsumoto et al., 2007; Rutkin, 2014) or “automated journalism” (Carlson, 2015; Lecompte, 2015; Napoli, 2014), “algorithmic journalism” (Anderson, 2013; Diakopoulos, 2014; Dörr, 2016), “machine-written

news” (Van Dalen, 2012; van der Kaa and Kraemer, 2014) or “computational journalism” (Anderson, 2011; Coddington, 2015, and this volume, Chapter 17; Flew et al., 2012; Gynnild, 2014; Karlsen and Stavelin, 2014; Stavelin, 2014), these are algorithm-driven automated processes using structured sets of data from sports, real estate, and stock markets as input to create news items as output. Platforms of advanced natural generation language transform data into text indistinguishable from what a human person would write (Clerwall, 2014; Dörr, 2016; Hammond, 2015; van der Kaa and Kraemer, 2014).

“Robot journalism” is a somewhat banal conceptualization, where illustrators often portray robots writing on computer keyboards instead of visually less attractive software. In this study, the unifying concept is computational journalism, which, according to Young and Hermida (2014: 381), refers to “forms of algorithmic, social scientific and mathematical processes and systems for the production of news” or, with a more normative approach, “the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism” (Hamilton and Turner, 2009: 2). We can simply conclude, like Coddington (2015), that computational journalism is concerned with the application of the processes of abstraction and automation to information. Here we are talking about intelligence augmentation (IA) rather than artificial intelligence (AI), following the path of the computer science pioneer Douglas Engelbart, who believed that computers should be used to “augment” or extend human capabilities rather than to mimic or replace them (Markoff, 2016: xii).

Journalists as a relevant group

This study departs from a social constructionist perspective, arguing that social action is a central element in technological development, that innovation in technology is a “multicentred, complex process”, and that there is a “spectrum of possible technological choices, alternatives and branching points” (Winner, 1993: 366). The theoretical focus is narrowed to social construction of technology (Bijker et al., 1987), and the fundamental argument is that human action shapes technology. The development and use of technology is embedded in its social context, in this case the newsroom.

Journalists are the main users of new technology in newsrooms, but following Winner (1993) begs questioning if and how the experiential knowledge of journalists is useful when it comes to the origin, character, and consequences of technical practices in a domain such as automated news production. To what extent are journalists considered a relevant social group that should be involved in development of automated news systems, or are they regarded as a brake on advancement? These questions are motivated by deep divisions in the debate on journalism and technological determinism. Studies of the introduction of new technology in newsrooms, from the computer onwards, show many examples of resistance, mitigation, and adaptation among journalists. Norms and conventions of journalism play an important role (Dickinson et al., 2013). Following Deuze (2005), it is understood that the ideology of journalism is a collection of “values, strategies and formal codes” that characterizes journalism and is shared by most members of the occupation as an active practice that is continually negotiated (Reese, 1990). Journalism ideology can be categorized in five different traits: journalists provide a public service and see themselves as watchdogs on behalf of a public; they are impartial, neutral, objective, and fair; they must be autonomous and free from bindings; they have a sense of immediacy, actuality, and speed; and journalists have a strong moral and ethical foundation that gives them a legitimate and valid reason to act on behalf of the public literature (Deuze, 2005; Golding and Elliott, 1999; Kovach and Rosenstiel, 2007).

From Goffman to artificial intelligence

Artificial intelligence is enjoying something of a renaissance after several false starts (Markoff, 2016). Increased computational resources, the explosive growth of digital data, a focus on specific

problems, knowledge engineering with machine learning, and alternative reasoning models have together contributed to the emergence of truly “intelligent” machines (Hammond, 2015). Many systems of artificial intelligence make use of computerized processes of human reasoning that sound familiar to any student of journalism, psychology, or sociology: what is happening around us (observing data), what does it mean (learning from it) and what is going to happen next? (Hammond, 2015; Hayes, 2013). Cognitive system engineering theories concentrate on human-like reasoning. In modeling human behavior, such as human practical reasoning, theories resonate well with, for instance, framing theory, which, as Carolyn Merchant (1989: xvi) notes, helps us to understand questions such as: ‘How do people “conceptualise” the world?’ ‘How do they “behave” in relation to the world?’ ‘How do they “give meaning” to the world?’ With Erving Goffman (1974), we navigate through these issues by asking two crucial questions that open the doors to framing theory as a method for understanding social reality. The first – “Under what circumstances do we think things are real?” – works to identify the conditions under which feelings of reality are established. The second – “What is it that is going on here?” – identifies the proper response to the current situation. In the same way computers are tuned to ask questions when trying to make sense of the environment and perform tasks, this is one-way journalists with limited understanding of computational thinking (Wing, 2011) can follow how reasoning and problem-solving is turned into software. Astrid Gynnild (2014: 724) actually asks “to what extent is training in logical, algorithmic, scientific thinking and in journalistic sorting and selection part of the same package?” Human-based and computer-driven approaches to sense-making and knowledge logic should be seen as complementary or supplementary (Flew et al., 2012). In other words, algorithmic and editorial logic should not compete in the newsroom bearing in mind that tightly categorized and structured information, the prerequisite for computerization, can at the same time have a damaging impact on human creativity (Kim and Zhong, 2017).

Human-machine advancement

The human-machine development in journalism began long before the introduction of the computer into newsrooms in the 1970s, and no empirical evidence suggests that there will be a break with the past, rather we can anticipate a continuous but rapidly accelerating technological development. Journalists are already surrounded by algorithms taking care of everything from web searches to photo and text editing, mostly using them unconsciously and seamlessly. Taking away these pieces of software would reveal to what extent journalists and editors already are dependent on automated decision systems. Algorithms are the result of human thinking and reasoning, but they are “neither entirely material, nor are they entirely human – they are hybrid, composed of both human intentionality and material obduracy” (Anderson, 2013: 1016). Nicholas Diakopoulos (2015; see also Koliska and Diakopoulos, this volume, Chapter 19) notes that human influences are embedded into algorithms, such as criteria choices, training data, semantics, and interpretation.

The principles of human-agent teaming in automation states that humans are responsible for outcomes in human-agent teams and therefore must be in command of the human-agent team (Billings, 1997; Tweedale et al., 2007; Urlings, 2004). To be in command, the human must be actively involved in the team process and, to remain involved, the human must be adequately informed. The human must be able to monitor agent behavior, and the activities of the agents must therefore be predictable. The agents must also be able to monitor performance of the human. To achieve all this, each team member (humans and agents) must have knowledge of the intent of the other members. In the computer science literature, agents are defined as computers with software: an agent can be seen as a software and/or hardware component of system capable of acting exactly in order to accomplish tasks on behalf of its user (Nwana, 1996).

The co-creator of the automated news provider Narrative Science, Kris Hammond, also a professor of computer science, underlines the importance of people in the workplace understanding how these new technologies work: “Black boxes that give us answers without explanations, or systems that fail to communicate with us, cannot become our trusted partners” (Hammond, 2015: 7). Therefore, people need to understand the basics of how these decision systems reason and come up with answers; otherwise, people cannot “anticipate the actions and limits of the automation” (Carr, 2015: 164).

Data and method

The empirical material consists of 35 qualitative research interviews with data journalists, news managers, computer scientists, academics, and industry experts in eight countries (US, UK, Sweden, Denmark, Norway, Germany, Spain, and France) that inhabit and reproduce this new ecosystem of computation, journalism, and statistics. The concept of saturation (Glaser and Strauss, 1967) has been the guiding principle: when the gathering of new data passes a point of diminishing return, it should be stopped – i.e., because it is too time consuming and not resulting in new information. Thus, this is not a representative randomized survey, but a study where the interviewees can be seen as key informants with special knowledge in the field, therefore also “participants” in the research. Rather than passively responding to questions, they were involved in a dialogue and have actively shaped the course of the interview (King, 1994). The duration of interviews was between 15 minutes and three hours, organized around a set of predetermined open-ended questions, with other questions emerging from the dialogue between interviewer and interviewee (DiCicco-Bloom and Crabtree, 2006). The choice of method was motivated by a need to focus on the meaning and experiential perceptions of automated news as well as historical accounts on an individual participant level. It is emphasized that qualitative research is concerned with meaning and not making generalized hypothesis statements (Mason, 2010). This study is exploratory work on an emerging field of automation and occupational effects where there is no broad agreement on the future position of journalism. Interviews were audio-recorded, transcribed (by transcription service Rev.com), and coded according to different aspects on the future of computational journalism. The transcripts were reread and compared over again until an interpretation of their meaning was obtained. As Steinar Kvale (2007: 109) notes, in principle such a hermeneutic text interpretation is an infinite process, whereas in practice it ends when a sensible coherent meaning has been reached.

Respondents were chosen based on their documented expertise in computational journalism (for instance, speakers at conferences, widely quoted in articles, blogs, and research reports) and were thought to be well placed to analyze the present and future of the industry. Snowball sampling (Goodman, 1961) was also used to find persons outside this group. These people represented a broad range of 25 different organizations such as new academic initiatives in digital journalism (Northwestern University, UNC School of Media and Journalism, University of Maryland, and University of Nebraska), local newspapers (Local Labs, Mittmedia), natural language generators of massive amounts of news texts (Automated Insights, Narrativa, AX-Semantics, Syllabs, Yseop, Narrative Science, Associated Press, United Robots), those analyzing official documents as a service (Nyhetsbyrå Sirén, Textual Relations), providing financial news (Thomson Reuters, Nasdaq, and Ritzau), and data journalists (at the *Guardian*, *Boston Globe*, The Lens, ProPublica, The Center for Investigative Reporting, and Journalism++ Stockholm).

The study addresses two research questions:

RQ1: What is the state of news automation in the United States and Europe?

RQ2: What are experts saying about constraints and opportunities for news automation?

Findings

Automated news was first introduced to newsrooms around 2009–2010 in the United States by software companies Narrative Science and Automated Insights, which have now established something of a duopoly. Their software already produces hundreds of thousands of texts for news organizations, including respected media companies such as Associated Press and Forbes.

Their automated news is dependent on tightly structured datasets for its operation. A dataset is a big table of numbers in which every row describes a single observation and every column describes a single attribute for each observation represented by the rows (Conway and White, 2012). The availability of data determines what sorts of texts can be generated. Automated Insights, for instance, relies on company earnings reports and analyst estimates to produce financial news for Associated Press, while Narrative Science uses data collected by team leaders with the mobile phone application Game Changer for sports news.

In Europe, automated news have also entered the commercial stage: companies such as AX-Semantics, Text-o-matic, Retresco, and Aexea in Germany, Narrativa in Spain, Syllabs in France, and Arria in the UK are offering limited services to media customers, mainly news on the stock market, weather forecasts, and football. In France, the interest in robot journalism has grown since Syllabs produced 34,000 articles for *Le Monde* during the departmental election night in March 2015. After this successful experiment, newsrooms in France have woken up to the possibility of producing articles on a massive scale and, according to a data journalist at Syllabs, “it is not difficult to get newsrooms to see the possibility with personalised content”.

In the Nordic region, several Swedish newspaper companies and one in Finland are publishing sports news from soccer games created by the service provider United Robots and based on match data collected by EverySport.com. A development manager at one of these companies, Mittmedia, which publishes 28 newspapers, noted that scalability of products is a challenge: “It is difficult to create generic solutions, we have to start from scratch for each new case, and relatively little is reusable”.

United Robots is looking for opportunities to automate publishing in cases where speed is important and content can be personalized. The company has started to extract news from large sets of data such as real estate transactions and earnings reports. Most studies of innovation in newsrooms, including automation, show that there is a “clear drive towards more efficient production processes” and the introduction of new technology is motivated by increasing productivity and saving costs (Dickinson et al., 2013). The fears of journalists are confirmed in the interviews: automated news is introduced to make newsrooms more efficient, for instance, by increasing the volumes. Automated news has been met with skepticism among the employees at Mittmedia (Svenlin, 2015), but managers say the purpose of introducing the new technology is not to get rid of employees and save costs but to create more local content and opportunities to sell local niche advertisement.

The news agency Siren is, with software developed by the consultancy Textual Relations, servicing media clients with automatically generated updates from tax authorities, administrative courts and the National Board for Consumer Disputes that can be used to write news reports. One of the developers at Textual Relations explains that for automated news either the input or the output has to be well-structured, “otherwise it becomes suddenly really complicated”. Another service provider, Journalism ++ Stockholm, creates automated texts for Siren’s customers based on, for instance, monthly crime statistics that are relevant to the specific location of the news agency’s customers.

In Denmark, the financial news agency Ritzau Finans has created templates for 70 listed companies to produce auto-flashes for earnings reports. The management believes that it makes sense investing in automation, not just because of speed but also for quality: the robot makes fewer

mistakes. They have also built into the system algorithms detecting potential errors, which means that earnings reports that divert too far from expectations are not auto-published. "It has taken some time to develop. We have to do a separate template for each company, but the job gets easier the more we do them", one Danish news manager claimed. After this first step, Ritzau has no further plans for automation, and progress so far seems to be quite modest.

There are also other newcomers in the Nordic region. In Norway, the news agency NTB has been producing soccer news automatically since 2016 with the help of software from NTB/Bakken & Bæk. In Finland, three separate election bots were developed for the local elections in April 2017, covering over 34,600 candidates and results from more than 300 different municipalities. One of these bots created news in three languages simultaneously, Finnish, Swedish, and English.

Social friction as barrier to automation

Social action is a central element in technological development. There are elements of friction that create barriers to increased automation in every form possible, ranging from culture and social norms to technical barriers. In news making, such friction might be an audience that is averse to automated news, another the ideology of journalism which has not lost its strength – much to the despair of a computer scientist interviewed for this study: "They're [journalists] very skeptical to the point of being [. . .] in a technology community they would be considered negative".

The risks are high that one core value of journalism, to question critically, challenge, and dispute, becomes a liability in the domain of automated news. Journalists are to an increasing extent dependent on colleagues with specialist skills in programming, while they also have a reputation for being bad at co-operating with colleagues with other skill sets. At the same time, journalists have been surprisingly resilient in refusing to trade in ideology for new knowledge, learning, and work opportunities (Boczkowski, 2004; Singer, 2004). However, recent examples of collaborative data journalism projects shows that sharing knowledge, tricks, and work practices is one of the core values of computational journalism. Basic technology literacy has become as important a skill as basic editorial skills (Anderson et al., 2012).

What to automate?

Whenever structured data are available, speed is essential, costs for automation investment go down, and the added value of the human touch in each particular piece of news is low, automation might take place. One perfect example would be financial news based on earnings reports, where speed is crucial, since the users of this information are basically trading software: the numeric data are stored inside SAP or other financial systems of the companies and are then summarized and put into a press release. Trading machines take the text and turn it back into numbers to make sense and turn the information into signals, buy, sell or hold. But the interviews show that neither data journalists nor news managers are sure if this will be a dominant feature of journalism. There is agreement around the notion that basic news stories just as well could be written by computers or, as a news manager said, "Writing a news story about what happened in a game yesterday? It is probably the most banal commodity for journalism I can think of.". This view is supported by a data journalist: "There are definitely a lot of routine, mundane items that computers can do quite competently, quite faster than humans, which free people up to work on other stories that computers can't". But this data journalist also believed news is too complex to be replaced by software in every part, and the same news manager continued: "Show me a news story that they generated algorithmically that isn't driven by a

highly structured, sequential, longitudinal dataset, like a sports story. Show me. And then I'll start worrying”.

There also seems to be some consensus around the notion that qualitative and experienced journalists will not be replaced, but they should focus on “higher value-added activities” (consultant). Asked under what circumstances the automation of journalism is worthwhile, a news manager in Denmark answers: “When you know that something is going to happen and you need to be fast. The next step is to anticipate unpredicted events”. Despite all the discussion about automated news and “robot journalism”, it should be noted that it is still a small domain within news making, with few signs of fast progress toward a future where automated news has become mainstream. Even technical experts of automated news are a bit skeptical regarding the added value. One Swedish developer says:

Personally, I have a hard time seeing value in texts written on a few data points. I think the real value in the long term will be to work with unstructured data sources such as alarm feeds and documents and use robots to identify news rather than to write it.

Distribution is one of the core issues. For a news agency like Associated Press or Ritzau, it makes sense to automate news because of the sheer volume of news and speed of operations. They also have an established ecosystem of services and customers: “If you want to create huge volumes of text you also need an organization which can receive it and distribute it,” a Swedish developer says.

Availability of data

Despite all the buzz about big data and the possibilities of mining huge sets of information, availability of data is actually one of the barriers to increased automation. One of the French journalists involved in the *Le Monde* project explained: “The difficulty in the election project was not in the writing of texts, it was more in collecting and cleaning data”, and continued, “the challenge is to get useful data, which is a problem in France”.

Data within the news production process is always a matter of choice, whether thinking about whom to interview or what set of statistics to use for a story. The integrity, quality, and reliability of the data available is crucial to automated news, just as it is to other forms of journalism. Therefore, the choice and evaluation of data should be a journalistic process, while validation, standardization, and normalization are normally handed over to programmers. A U.S. data journalist explained that getting interesting and useful data is often the result of long struggles based on Freedom of Information Act (FOIA) requests. He has observed a clash within government agencies between people “who really believe that the data belongs to the people” and that data should be published openly, on the one hand, and other people “who feel very proprietary and territorial about it, and want to keep everything in-house and worry about releasing it to the world”. Even though government agencies are publishing more data online, he still has to fight hard to get other data that the same agencies keep. Another U.S. data journalist confirmed that the information she uses is “almost always” the result of a FOIA request. Open access is no guarantee, either: in Sweden, with its 250-year history of having the world's first FOIA, protocols from meetings in all 290 municipalities are available online. Data journalism agency Journalism++ Stockholm has scraped them all, and the work reveals that “there are probably more ways of writing protocols than there are municipalities”, according to one of the project managers. Thus, the availability of data that are both structured and interesting becomes a constraint.

Automation for predictive journalism

Automation of news making can address managerial concerns to develop more efficient editorial processes and a turn toward an anticipatory journalism, focused more on predictive than reactive routines to use resources more efficiently. News alarms like those based on automated searches of police arrest records that *Los Angeles Times* crime reporters are using will be much more common. Local Labs, a company in Chicago, is one example of how automation for proactive or predictive news making is used in a coherent way. Local Labs, according to its web page, provides “community news production, advertising and advertorial content, print special sections and web and mobile application management” (www.locallabs.com). The company is using a combination of a database with sources, a calendar, and scheduled email to automatically collect updates on activities in local communities. Certain events tend to get repeated with some frequency, like activities related to Christmas, and a news manager at the company explains: “There are certain things that require reactive resources. The reactive resources should be not the norm”. Preexisting information – events that the newsroom has knowledge on – is used to predict events through a practical application of probability theory. This saves the journalists time and effort and makes news making more efficient. A U.S. data journalist believed prediction is probably one of the most useful applications of news automation:

There are a lot of things that are predictable, where you know likely what form that information's going to come in, so you could write a program that could interpret that predictable feed and write a predictable story based on it. It's harder for things that are more nuanced, or less predictable.

Limited role of journalists

The data journalists interviewed for this study expressed feelings of both acceptance and criticism of automated news. That is understandable for several reasons. All innovation involves unintended and undesirable consequences, which are seldom discussed in the innovation research literature (Sveiby et al., 2009). Journalists have a legitimate reason to think about new technology critically and discuss how it affects their work. Automated news has been introduced by software companies and sold as a service or plug-and-play software. The role of journalists in developing automated news has been limited to “training” the algorithms to choose the proper wording for natural language generation, and in some cases where journalists have been involved, the experience has not been a pleasant one.

Journalists, moreover, are not ready to accept machine-written texts based on a dozen different templates except for basic reports on sports or the weather. A data journalist explained why her brief experience with automation was interesting but not something she wanted to repeat. She called it a “nightmare” because of the work involved in checking computer-generated language: “When we write a story with numbers, we painfully pick the right words to say each number and report about it”. This statement stands in stark contrast to another by a computer scientist: “a lot of the training in journalism is really more about finding the story and finding the information and less about the writing per se”.

Journalists are probably right to fear the risk that templates and formulas for automated news are reductive in terms of news diversity and complexity. Computer intelligence is predictive and entirely lacking in curiosity, imagination, and worldliness (Carr, 2015: 124). To be objective, fair, and credible is a key element in the self-perception of journalists, and using appropriate words is an expression of this: choosing the right vocabulary that does not misrepresent the information in

numbers is not a minor issue. There is also concern with getting facts right, as one U.S. data journalist said: “Fast is good. Accuracy is more important, so those two items are always competing”.

Only the beginning

NLG technology is only gradually being developed for more sophisticated storytelling (Caswell and Dörr, 2017), and there is reason to believe that present news automation practices are just a taste of the future. The first wave of news written by software will be followed by much more advanced systems that not just journalists but also programmers will have difficulties to supervise: Machine learning means that “computer programs build their own algorithms through training, feedback, and iteration, and are therefore much more difficult to understand and control” (Bogen, 2015). Even when relatively simple NLG techniques are implemented in editorial environments, the complexity increases very quickly for most useful applications (Caswell and Dörr, 2017).

In the existing model for automated news, algorithms are precise, perfect, meticulously programmed and severe in manner. Human engineers are giving instructions to the computer that will handle certain inputs and generate certain outputs: the logic of the program is the code written by humans. Machine learning is about building programs that build themselves and are, for instance, used to predict users’ preferences by companies such as Netflix and Amazon. These machine-generated programs are “trained” by their designers through giving feedback on the results. “They are difficult (sometimes impossible) to understand, tricky to debug, and harder to control. Yet it is precisely for these reasons that they offer the potential for far more ‘intelligent’ behavior than traditional approaches to algorithms and A.I.” (Auerbach, 2015).

These are learning algorithms that make other algorithms figure it out on their own, by making inferences from data: “The more data they have, the better they get” (Domingos, 2015: xi). This alternative model of programming and computation “sidesteps the limitations of the classic model, embracing uncertainty, variability, self-correction, and overall messiness” (Auerbach, 2015). Translated to journalism it means that computer programs will not need to be instructed to understand the complex, often subconscious and contradictory processes of journalistic work to produce news. Instead, using a corpus of news texts, they will do the job. “If we give a learner [a learning algorithm] a sufficient number of examples of each, however, it will happily figure out how to do them on its own” (Domingos, 2015: 6).

The risk is that media companies will not be able to face this technological challenge. One European provider of NLG solutions says he negotiated for three months with a media company, which ultimately said no to the offer because it was deemed too expensive: “Never again will we create any more prototypes before we see the money”. A representative for an American service provider expresses similar frustration. “Media companies are difficult customers to deal with and they have no money”. Instead, this company is focusing on the financial industry. A Swedish representative for a service provider says that the news business is good for testing systems, but his company is planning to service customers outside the media industry instead. One of the most promising customer segments are companies involved in global e-commerce, as they have a great need for updating product information in multiple languages on their websites. Therefore, the risk is obvious that media companies for a number of reasons will fall behind other industries that use cognitive technologies.

Conclusion

This exploratory study turns the usual discussion about job losses on its head with a claim that “algorithms are a journalist’s new best friend”. The real risk is not too much automation but that

journalists and the news media are unable to follow other industries in developing new cognitive technologies because of, for instance, lack of skills, strategy, and scarce financial resources. In reality, the latest development – popularly but somewhat misleadingly named “robot journalism” – is just another step in human-computer advancement and cognitive technologies that simulate human reasoning and perceptual skills. I argue that besides the general conclusion that creative jobs such as journalism are at low or no risk to automation, journalism as ideology, understanding how journalists give meaning to their work (Deuze, 2005), probably also will be a strong mitigating effect in the future. Journalists are working in a “complex and contradictory set of macro-sociological influences” and make a good case study to understand “occupational discourses in times of rapid economic and cultural change and widespread professional delegitimation” (Anderson, 2014). The discussion on why there still are so many jobs in journalism could be relevant to broader arguments about resilience in creative and artisanal jobs and also would deserve more interest outside the field of journalism studies or media management research.

Further reading

This chapter has benefitted tremendously from ground-breaking work on sociology of journalism by Mark Deuze, especially the 2005 article “What is Journalism? Professional Identity and Ideology of Journalists Reconsidered”. Regarding emerging forms of newswork and professional journalism, C.W. Anderson’s work has been very helpful, including “Towards a Sociology of Computational and Algorithmic Journalism” (2013). A great source of inspiration are the writings of Nick Diakopoulos, especially the 2014 report *Algorithmic Accountability Reporting: On the Investigation of Black Boxes* for the Tow Center for Digital Journalism, Columbia School of Journalism.

References

- Akst, D. (2013) “Automation anxiety.” *The Wilson Quarterly Summer*. Retrieved from <http://archive.wilsonquarterly.com/sites/default/files/articles/AutomationAnxiety.pdf>
- Anderson, C. W. (2011) “Notes towards an analysis of computational journalism.” *HIIG Discussion Paper Series*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2009292 (accessed 11 January 2016).
- Anderson, C. W. (2013) “Towards a sociology of computational and algorithmic journalism.” *New Media & Society*, 15(7), 1005–1021.
- Anderson, C. W. (2014) “The sociology of the professions and the problem of journalism education.” *Radical Teacher*, 99, 62–68.
- Anderson, C. W., Bell, E. and Shirky, C. (2012) *Post-industrial Journalism: Adapting to the Present*. New York, NY: Tow Center for Digital Journalism, Columbia University.
- Atkinson, R. D. and Wu, J. (2017) *False Alarmism: Technological Disruption and the U.S. Labor Market, 1850–2015*. Washington: Information technology & Innovation foundation.
- Auerbach, D. (2015) “The programs that become the programmers.” *Slate, New America and ASU*. Retrieved from www.slate.com/articles/technology/bitwise/2015/09/pedro_domingos_master_algorithm_how_machine_learning_is_resaping_how_we.html (accessed 11 December 2015).
- Autor, D. (2014) *Polanyi's Paradox and the Shape of Employment Growth*. Cambridge, MA: National Bureau of Economic Research.
- Bakhshi, H., Frey, C. B. and Osborne, M. (2015) *Creativity Vs. Robots: The Creative Economy and the Future of Employment*. London: Nesta.
- Bijker, W., Hughes, T. and Pinch, T. (eds.). (1987) *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology* (Vol. 17). Cambridge, MA: The MIT Press.
- Billings, C. E. (1997) *Aviation Automation: The Search for a Human-Centered Approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Boczkowski, P. J. (2004) “The processes of adopting multimedia and interactivity in three online newsrooms.” *Journal of Communication*, 54(2), 197–213.

- Boczkowski, P. J. (2005) *Digitizing the News: Innovation in Online Newspapers*. Cambridge, MA: The MIT Press.
- Bogen, M. (2015) "Algorithms of war." *Slate, New America and ASU*. Retrieved from www.slate.com/articles/technology/future_tense/2015/12/the_dangers_of_enlisting_algorithms_in_statecraft.html?wpsrc=sh_all_tab_tw_ru (accessed 11 December 2015).
- Bureau of Labor Statistics. (2016–2017 Edition) *Occupational Outlook Handbook: Reporters, Correspondents, and Broadcast News Analysts*. Washington, DC: U.S. Department of Labor.
- Carlson, M. (2015) "The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority." *Digital Journalism*, 3(3), 416–431.
- Carr, N. (2015) *The Glass Cage: Where Automation Is Taking Us*. London: The Bodley Head.
- Caswell, D. and Dörr, K. (2017) "Automated journalism 2.0: Event-driven narratives: From simple descriptions to real stories." *Journalism Practice*, 1–20.
- Clerwall, C. (2014) "Enter the robot journalist." *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.
- Conway, D. and White, J. M. (2012) *Machine Learning for Hackers*. Sebastopol, CA: O'Reilly Media.
- Deuze, M. (2005) "What is journalism? Professional identity and ideology of journalists reconsidered." *Journalism*, 6(4), 442–464.
- Diakopoulos, N. (2014) *Algorithmic Accountability Reporting: On the Investigation of Black Boxes*. New York, NY: Tow Center for Digital Journalism, Columbia School of Journalism.
- Diakopoulos, N. (2015) "Algorithmic accountability: Journalistic investigation of computational power structures." *Digital Journalism*, 3(3), 398–415.
- DiCicco-Bloom, B. and Crabtree, B. F. (2006) "The qualitative research interview." *Medical Education*, 40(4), 314–321.
- Dickinson, R., Matthews, J. and Saltzis, K. (2013) "Studying journalists in changing times understanding news work as socially situated practice." *International Communication Gazette*, 75(1), 3–18.
- Domingos, P. (2015) *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*. St Ives: Allen Lane.
- Dörr, K. N. (2016) "Mapping the field of algorithmic journalism." *Digital Journalism*, 4(6), 700–722.
- Flew, T., Spurgeon, C., Daniel, A. and Swift, A. (2012) "The promise of computational journalism." *Journalism Practice*, 6(2), 157–171.
- Glahn, H. R. (1970) "Computer-produced worded forecasts." *Bulletin of the American Meteorological Society*, 51(12), 1126–1131.
- Glaser, B. and Strauss, A. L. (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York, NY: Aldine Publishing Company.
- Goffman, E. (1974) *Frame Analysis: An Essay on the Organization of Experience*. New York, NY: The Maple Press.
- Golding, P. and Elliott, P. (1999) "Making the news." In H. Tumber (ed.), *News: A Reader*. Oxford: Oxford University Press (pp. 112–120).
- Goodman, L. A. (1961) "Snowball sampling." *The Annals of Mathematical Statistics*, 32(1), 148–170.
- Gynnild, A. (2014) "Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets." *Journalism*, 15(6), 713–730.
- Hamilton, J. T. and Turner, T. (2009) "Accountability through algorithm: Developing the field of computational journalism." *Center for Advanced Study in the Behavioral Sciences*. Retrieved from <http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>
- Hammond, K. (2015) *Practical Artificial Intelligence for Dummies*. Hoboken, NJ: John Wiley & Sons.
- Hargreaves, I. (2016) "Foreword." In N. Thurman, A. Cornia and J. Kunert (eds.), *Journalists in the UK*. Oxford: Reuters Institute for the Study of Journalism.
- Hayes, A. F. (2013) *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford Press.
- Karlsen, J. and Stavelin, E. (2014) "Computational journalism in Norwegian newsrooms." *Journalism Practice*, 8(1), 34–48.
- Kim, Y. J. and Zhong, C.-B. (2017) "Ideas rise from chaos: Information structure and creativity." *Organizational Behavior and Human Decision Processes*, 138, 15–27.
- King, N. (1994) "The qualitative research interview." In C. Cassel and G. Symon (eds.), *Qualitative Methods in Organizational Research: A Practical Guide*. London: Sage Publications (pp. 1–35).
- Kolkey, D. (1999) "Features-newsroom automation-new systems offer efficient solutions to newsroom staffs." *Broadcast Engineering*, 41(10), 94–99.
- Kovach, B. and Rosenstiel, T. (2007) *The Elements of Journalism: What Newspeople Should Know and What the Public Should Expect*. New York, NY: Three Rivers Press.

- Kvale, S. (2007) *Doing Interviews: The Sage Qualitative Research Kit*. Los Angeles, CA: Sage Publications.
- Latar, N. L. (2014) "Robot journalists: 2015. 'Quakebot' is just the beginning." *Wharton, University of Pennsylvania*. Retrieved from <http://knowledge.wharton.upenn.edu/article/will-robot-journalists-replace-human-ones/> (accessed 15 January 2015).
- Lecompte, C. (2015) "Automation in the newsroom." *Nieman Reports*, 69(3), 32–45.
- Leppanen, L., Munezero, M., Granroth-Wilding, M. and Toivonen, H. (2017) "Data-driven news generation for automated journalism." *International Conference on Natural Language Generation, INLG 2017*, 4–7 September 2017, Special Interest Group on Natural Language Generation (SIGGEN).
- Levy, S. (2012) "The rise of the robot reporter." *Wired*, 20(5), 132–139.
- Linden, C.-G. (2017) "Algorithms for journalism: The future of news work." *The Journal of Media Innovations*, 4(1), 60–76.
- Marconi, F. and Siegma, A. (2017) *The Future of Augmented Journalism: A Guide for Newsrooms in the Age of Smart Machine*. New York, NY: Associated Press.
- Markoff, J. (2016) *Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots*. New York, NY: Harper Collins.
- Mason, M. (2010) "Sample size and saturation in PhD studies using qualitative interviews." *Forum: Qualitative Social Research*, 11(3), 1–19.
- Matsumoto, R., Nakayama, H., Harada, T. and Kuniyoshi, Y. (2007) "Journalist robot: Robot system making news articles from real world." *IEEE*, April 10–14, 2007.
- Meehan, J. R. (1977) "Tale-spin, an interactive program that writes stories." Paper presented at the *IJCAI'77 Proceedings of the 5th International Joint Conference on Artificial Intelligence* (pp. 91–98).
- Merchant, C. (1989) *The Death of Nature: Women, Ecology, and the Scientific Revolution*. San Francisco, CA: Harper and Row.
- Napoli, P. M. (2014) "Automated media: An institutional theory perspective on algorithmic media production and consumption." *Communication Theory*, 24(3), 340–360.
- Nwana, H. S. (1996) "Software agents: An overview." *Knowledge Engineering Review*, 11(3), 205–244.
- Office for National Statistics. (2016) *EMP04: Employment by Occupation*. Retrieved from www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/employment-byoccupationemp04 (accessed 6 May 2017).
- Pavlik, J. (2000) "The impact of technology on journalism." *Journalism Studies*, 1(2), 229–237.
- Reese, S. D. (1990) "The news paradigm and the ideology of objectivity: A socialist at the wall street journal." *Critical Studies in Media Communication*, 7(4), 390–409.
- Rutkin, A. (2014) "Rise of robot reporters: When software writes the news." *New Scientist*, 221(2962), 22.
- Scherer, E. (2010) "Le 'Journalisme Augmenté' En 10 Points." *OWNI*. <http://owni.fr/2010/11/07/le-%C2%AB-journalisme-augmente-%C2%BB-en-10-points/> (accessed 15 May 2017).
- Shipman, J. M. (1986) "Computerization and job satisfaction in the newsroom: Four factors to consider." *Newspaper Research Journal*, 8(1), 69–78.
- Singer, J. B. (2004) "Strange bedfellows? The diffusion of convergence in four news organizations." *Journalism Studies*, 5(1), 3–18.
- Smith, A. and Anderson, J. (2014) *AI, Robotics, and the Future of Jobs*. Washington, DC: Pew Research Center.
- Stavelin, E. (2014) *Computational Journalism: When Journalism Meets Programming*. PhD thesis, The University of Bergen.
- Sveiby, K.-E., Gripenberg, P., Segercrantz, B., Eriksson, A. and Aminoff, A. (2009) "Unintended and undesirable consequences of innovation." *ISPIM*, 21–24 June 2009.
- Svenlin, Johan. (2015) "Roboten Kommer!" *Journalisti*. Accessed December 16, 2015. www.journalisti.fi/artikkelit/2015/15/roboten-kommer-/
- Taibi, C. (2014). "Should we be afraid or excited about robot journalism?" *Huffington Post*. Retrieved from www.huffingtonpost.com/2014/07/11/robotjournalism-good-technology-_n_5577651.html (accessed 23 December 2015)
- Tornøe, R. (2014) "Learn to stop worrying and love robot journalists." *Editor & Publisher*, 147(9), 24–26.
- Tweeddale, J., Ichalkaranje, N., Sioutis, C., Jarvis, B., Consoli, A. and Phillips-Wren, G. E. (2007) "Innovations in multi-agent systems." *Journal of Network and Computer Applications*, 30(3), 1089–1115.
- Urlings, P. (2004) *Human – Machine Teaming*. PhD thesis, University of South Australia.
- 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. and Kraemer, E. (2014) "Journalist versus news consumer: The perceived credibility of machine written news." *Columbia University*, 24–25 October. http://compute-cuj.org/cj-2014/cj2014_session4_paper2.pdf.

- Vigneaux, S. (1996) "The integration of a newsroom computer system with a server-centered news production system." *IET*. Retrieved from <https://ieeexplore.ieee.org/abstract/document/643225/>
- Weaver, D. H. (2015) "Studying journalists and journalism across four decades: A sociology of occupations approach." *Mass Communication and Society*, 18(1), 4–16.
- Wing, J. (2011) "Research Notebook: Computational Thinking—What and Why?" Retrieved from <https://www.cs.cmu.edu/link/research-notebook-computational-thinking-what-and-why>
- Winner, L. (1993) "Upon opening the black box and finding it empty: Social constructivism and the philosophy of technology." *Science, Technology, and Human Values*, 18(3), 362–378.
- Young, M. L. and Hermida, A. (2014) "From Mr. and Mrs. Outlier to central tendencies." *Digital Journalism*, 3(3), 381–397.