Algorithms for journalism: The future of news work

Carl-Gustav Lindén

University of Helsinki

carlgustavl@gmail.com

ABSTRACT

Software-generated news, sometimes called "robot journalism," has recently given rise to concerns that the automation of news will make journalists redundant. These arguments follow a deterministic line of thinking. Algorithms choose information for users but are also the construct of social process and practice. The aim of this essay is to explore "the algorithmic turn" (Napoli, 2014a) in news production. Based on case studies from three separate news outlets it is found that the impact of automated news is, first, increased efficiency and job satisfaction with automation of monotonous and error-prone routine tasks; second, automation of journalism routine tasks resulting in losses of journalist jobs; and third, new forms of work that require computational thinking.

INTRODUCTION

This essay is based on an exploratory study of a recent development in computational journalism: news generated by software where tens of thousands of similar stories are automatically written based on large datasets. The study draws upon qualitative data, a diversity of opinions expressed in 24 exploratory expert interviews. The aim is to add to a limited body of research on the impact of algorithms on the production of media content as well as how this is disrupting journalism practice and the normative foundation of journalism. Re-

cently, software-generated news – somewhat misleadingly called "robot journalism" – has risen to the headlines in accounts of how the automation of news will make journalists redundant. These arguments follow a deterministic logic that rings familiar to sociologists researching technology, work and organisations; yet the capacity and effects of technology are essentially conditional to social construction, not a technically transparent description (Grint & Woolgar, 1997, p. 165). Algorithms can be likened to a knowledge machine that chooses information for users, but also to a social process by which these formulas are made legitimate in the system (Gillespie, 2014). The aim of this essay is to explore "the algorithmic turn" (Napoli, 2014a) in existing operations of automated news from an in-

Keywords

stitutional perspective on technology and journalism work.

Two research questions were asked. RQ1: How do managers, journalists, and programmers working with news innovations anticipate the future of software-generated news? What do they think about the role of journalists? RQ2: How can automation help elevate journalists from repetitive routine tasks to more advanced forms of reporting?

Three separate cases of existing news automation are presented. The analysis is inspired by a discursive framework for the relationship between journalistic work and technology developed by Matthew Powers (2012). The impact of algorithms can be 1) increased efficiency and job satisfaction with automation of monotonous and error-prone routine tasks, 2) automation of journalism routine tasks resulting in losses of journalist jobs, and 3) new forms of work "yet-to-be-invented" that require computational thinking.

TECHNOLOGY AND JOURNALISM

Technology – the combination of technical devices and instructions to make them work, such as software – has during the centuries been one of the key factors in the development of journalism and media (McLuhan, 2011; Örnebring, 2010; Pavlik, 2000; Powers, 2012) and rapid technological changes force journalists to upgrade their skills (Örnebring, 2016). Consider a time quite recently when newsrooms had no access to photocopying machines, mobile phones, the Internet, email or digital equipment such as cameras or voice recorders. Imagine how these new opportunities have changed the way journalists interact with the world outside the newsroom, where at the same time processual operating tools have made it easier to manage editorial tasks, such as editing, proof reading, visualisation, and content design.

When Philip Meyer worked on his eminent book *Precision Journalism* in the 1960s he wanted to introduce scientific methods from social sciences to journalists. We have come a long way from the first version of the book (P. Meyer, 1973) where his main tool for making advanced use of data was the slide rule. By the time the next edition was published in 1979 Meyer was already instructing journalists how to use computers. Still, software was rudimentary, processing power weak and the computer had yet to arrive on a massive scale in newsrooms. The first pioneers had to write their own software, or ask coders to do the job. Fifteen to twenty years later, computers had "near complete adoption" in newsrooms (Garrison, 2001a) though used on an advanced level by a few (Maier, 2000).

The speed of technological change is maybe higher than in any previous instance in the history of journalism. The diffusion of technology in the world is accelerating: while it took 75 years to reach fifty million users of telephones, the digital game Angry Birds collected the same number of users in 35 days. And a service like WhatsApp gained more followers during its first six years in existence than Christianity in its first nineteen centuries (Frey & Osborne, 2015). However, with diffusion of new technology comes tension in the newsroom: the division of labour between "traditional" journalists and colleagues more adept at using new technology seems to widen in times of change until the early adopters are joined by others (Garrison, 2001b; Maier, 2000; Singer, 2004). For instance, while multimedia news is already well established, a larger work force of multimedia journalists has been slow to arrive to the newsroom (Hermida & Thurman, 2008; Saltzis & Dickinson, 2008).

Now, journalists are facing an even greater transformation: the automation of journalism work (Clerwall, 2014; Kim, Lee, Kim, Kuppuswamy, & Jo, 2007; Latar, 2014; Matsumoto, Nakaya-

ma, Harada, & Kuniyoshi, 2007; Napoli, 2014b; Van Dalen, 2012), such as the use of algorithms as a way to gain insights about what is engaging the audience (Edge, 2014). The technology in this case consists of newsroom computerisation in the form of software-generated news – a concept somewhat overlapping other terms such as artificial intelligence (AI) news, robot journalism, bot-driven or data-driven journalism, computational journalism or algorithmic journalism – and it is already having an impact on journalism practice (Karlsen & Stavelin, 2014; Stavelin, 2014).

COMPUTATIONAL JOURNALISM

I consider computational journalism the most proper unifying concept, which, according to Young and Hermida (2014), 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 & Turner, 2009, 2). This implies a systemic change in the way "stories

are discovered, presented, aggregated, monetised, and archived" (Cohen, Hamilton, & Turner, 2011). In our case news "robots" are computer programs containing algorithms that detail the specific instructions a computer should perform (in a specific order) to replicate the end results of journalism.

Algorithms do not have a generally accepted formal definition, but the term often 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 precisely defines a sequence of precise instructions that can be understood by a computer. Sometimes the question is asked: Is there an algorithm for journalism? The answer is yes, but to a certain degree. An algorithm can also contain instructions for people who work with routine tasks in journalism. "For people to follow the rules of an algorithm, the rules must be formulated so that they can be followed in a robot-like manner, that is, without the need for thought" (Stone, 1971, pp. 4-5).

In that case the instructions would be written out in "pseudocode," a non-computer, non-compilable language that is computer-like but still contains enough natural language to be readable (Berry, 2011). In plain English it would read something like:

Find a piece of new information that adheres to a fixed set of news rules such as human curiosity, relevance and impact, contact three publicly recognised and independent sources for comment and context, produce an article of 200 words written according to the newsroom stylebook, and submit that to an editor or directly to the audience.

These basic instructions can be followed by computers as well and the concept of a strictly rule-based form of journalism work is applied in automated news, however only in a very limited domain and I will come back to that. What about journalism higher up in the knowledge hierarchy – can that be simulated and optimised by smart machines? In principle, an institutional approach to journalism would depart from the understanding that news is the result of adaption of certain normative obligations that "come to take a rule like status in thought and action" (J. W. Meyer & Rowan, 1977, p. 341) produced inside a physical organisation or a cultural and social system.

JOURNALISM AS TACIT KNOWLEDGE

In most countries, practicing journalism has never depended on a systematic body of knowledge acquired through formal training (Powers, 2012). It is also known that journalists are not very confident of their occupational roles (Sparrow, 1999; Tunstall, 1975) and important features of their work, such as passion or creativity, are mostly left out of an equation that serves to support claims to professionalism. Journalism is, to a large extent, based on what Michael Polanyi called tacit knowledge, meaning "we can know more than we can tell" (Polanyi, 1967, 4). This matter is illuminated, for instance, by the way journalists try to explain the most decisive rules in journalism: the newsworthiness rules. Rules can be regarded as routines, procedures, conventions, roles, strategies, organisational forms, and technologies around which a certain activity is constructed (March & Olsen, 1989). Around these roles and routines there is a web of "beliefs, paradigms, codes, cultures, and knowledge" that both support and contradict them. Thus, these procedural rules in the newsroom in theory stipulate what kind of events become a story, what facts are relevant, the appropriate setting in time for that report, who has the authority to be interviewed for a story, and how that story shall be presented. However, these rules are changing with conditions and are often inconsistent, thus they are not automatic.

The discrepancy between the news criteria and the way news turns out in practice has been something of a mystery to journalism researchers. News should be accurate, fresh, timely and unexpected, both interesting and important, but as Timothy Cook (1998) notes, beyond that nobody knows: "Reporters, asked what newsworthiness is, usually cannot answer the questions" (p. 72). One reason could be that daily news work is grounded in the production processes (Ess, 2014; Thorén, 2014) and does not offer much opportunity for paralysing self-reflection. An alternative answer would be that journalism actually is more creative and improvisational than researchers tend to think, despite a certain reliance on routines and rules. Even though there are occupations where persons acquire skills and reach the "automated phase of learning" (Ericsson, 2004, p. S70) this cannot be the case in journalism confronted by uncertainty where continued deliberate practice is necessary for maintenance of skills performance.

News, for instance, is often a product of negotiation and social engagement with sources where the "rules of appropriateness" (March & Olsen, 1989) can be quite different from formal rules that might serve as rituals, for instance celebrating the watchdog role of journalists. I would also add that breaking these rules of appropriateness can be necessary in cases when information is withheld and sources are hard to get. Further, the influence from the institutional environment should not be neglected since comparative studies show that similar normative obligations are translated into role perceptions that diverge in different countries (Hanitzsch et al., 2011).

At least in normative journalism theory, if not always in popular discourse, journalists are what Rushworth Kidder (1995) defines as "good people"; they "seem to have some conscious sense of vision, some deep core of ethical values that gives them the courage to stand up to the tough choices." The external moral guidance such as institutional norms and principles typified in codes of ethics journalists are supported by an internal view of the virtues of justice and integrity, a "moral psychology for journalists based in virtue" (Quinn, 2007).

Thus a journalism algorithm needs to reflect a behaviour that is not always rule-based but reflective and flexible in the sense that the application of explicit and internalised rules to complex situa-

tions is ambiguous. Determining what is appropriate in a specific situation is a "nontrivial exercise" (March & Olsen, 1989, p. 25).

AUTOMATION AND JOURNALISM

In spite of this uncertainty about the rule-based part of journalism, codification of journalism knowledge is already happening; specific forms of news work have been broken down to parts that can be automated. Development in rudimentary AI for journalism and media is fast and commercial applications of automated news services provided by, for instance, the two American companies Narrative Science and Automated Insights, are producing millions of stories based on structured data from financial markets and sports events for media customers such as The Big Ten Network, Forbes, and the Associated Press. Computers are programmed to produce standard stories that are almost impossible to distinguish from articles written by humans (Clerwall, 2014) – and algorithms are trained to become smarter. It is worth noting that Facebook has been accused of employing journalists to train its algorithms to select stories for the news feed that might interest people (Nunez, 2016).

QUANTITATIVE OPPORTUNITIES

The rapid development of software-generated news coincides with a quantitative and computational shift in journalism where reporters are feeling more confident in using statistical methods for mining and analysing data and creating stories based on this material (Anderson, 2011; Coddington, 2015; Cohen, Li, Yang, & Yu, 2011; Flew, Spurgeon, Daniel, & Swift, 2012; Howard, 2014; Parasie & Dagiral, 2012). For instance, machine learning techniques have become an integral part of editorial work at big media companies such as the Guardian, Reuters, and the New York Times. Traditionally, journalists have mainly relied on qualitative methods for their research, but the supply of easy-to-use, cheap, or free software for processing large data sets has opened new opportunities. The personal memory capacity and analytical processing capabilities of journalists have already been augmented with the aid of computers.

The digital revolution has also expanded the supply and availability of data that can be used for computational journalistic processes; this is the explosion of what is popularly labelled "Big Data" (Fairfield & Shtein, 2014; Lewis & Westlund, 2014). Journalists are transferring from a world where they are exposed and reacting to largely unstructured information into a structured information environment where events can be anticipated to a larger extent than before.

AUTOMATION ANXIETY

"Robot journalism" has in the popular discourse been regarded both as a threat and a savior for human journalists. Automation or computer anxiety is certainly not a new thing in either knowledge work in general or journalism in particular. Aristotle, Queen Elisabeth I, the Luddites, James Joyce and John Maynard Keynes were all concerned with the impact of technology on employment (Acemoglu & Robinson, 2012; Akst, 2013).The advent of the computer in newsrooms some thirty years ago created tension and journalists are actually the only ones who have survived the change while other non-editorial occupations such as typesetters, telephone operators, and darkroom assistants have all but disappeared (Linden, 2017).

The digital revolution has mainly benefitted people as consumers, creating many things offered for free, including news. On the other hand, new employment opportunities have mainly been cre-

ated for highly skilled workers. In the digital age, innovators and entrepreneurs are the main beneficiaries. Oxford researchers (Frey & Osborne, 2013) predict that 45% of American occupations will be automated within the next 20 years. The first stage will be using computational power to replace jobs that rely on such things as pattern recognition, data gathering and distillation, and computational algorithms. The researchers also predict that artificial intelligence will eventually put jobs in management, science, engineering, and the arts at risk.

Further, even though creative jobs such as journalism in general are likely not to be automated, forms of media such as newspapers are more susceptible to automation (Bakhshi, Frey, & Osbourne, 2015), and certain routine tasks that journalists perform will continue to be automated as well. Research does not point to automation as a cause of direct job losses in creative occupations, but the impact can be felt in many indirect ways. In conclusion, there is an algorithm for journalism, but even though some parts of news work – "low journalism" – will be automated, there is reason to believe that more demanding forms – "high journalism" – will benefit from man-machine combinations that ideally would have the journalist in charge.

METHOD AND DATA

This journalism-centric or "internalist" (Anderson, 2013) organisational-level case study explores current software-generated news practices, mainly in the United States, from a systemic occupational perspective. News automation has reached the market stage in the U.S. while still being in the pre-market stage in much of Europe (Dörr, 2015). The study draws upon qualitative data, a diversity of opinions expressed in 24 exploratory expert interviews. The initial respondents were identified and interviewed at three separate conferences about data journalism. These were the Global Editors Network Summit in Barcelona, Spain in June 2014, the Nordic Data Journalism Conference (NODA15) in Ålesund, Norway, in January 2015 and the National Institute for Computer-Assisted Reporting (NICAR) conference in Atlanta, USA in March 2015.

In addition to this, expert interviews were conducted in person at two universities that combine education and research in journalism with computer science, as well as three commercial companies and one non-profit organisation that have experience in developing and distributing automatically generated news. Three cases were selected based on information gained in these interviews and from industry reports. These cases were the Associated Press and ProPublica in New York and Local Labs in Chicago; from these organisations, more focused interviews with a total of five key people were then conducted.

The data was obtained through dialogical interviews which provided the researcher with descriptions, narratives, and texts (Kvale, 2005). The discussions were recorded and transcribed before interpretation of meaning. Of the total number of interviewees from the conferences and the news organisations, eight were data journalists, three news editors, two managers of news operations, six scholars with an interest in computational journalism, four representatives of companies producing news automation services, and one was a tech expert in the financial markets. Out of these 24 people, only four were women, which reflects that computational journalism is mainly populated by men, rather than the result of a biased selection of people to interview.

As a method, expert interviews are particularly appropriate in research aimed at reconstructing explicit expert knowledge (Pfadenhauer, 2009). Talking to experts is an especially efficient method of gathering data in the exploratory phase of a

research project compared with, for instance, participatory observation or systematic quantitative surveys (Bogner, Littig, & Menz, 2009). Experts are people who hold key positions in their organisations and in the case of this study their support, possibly motivated by a shared understanding of the social relevance of the research, made it easier to gain access to an extended circle of experts.

Two research questions were asked:

*RQ*1: How do managers, journalists, and programmers working with news innovations anticipate the future of software-generated news? What do they think about the role of journalists?

RQ2: How can automation help elevate journalists from repetitive routine tasks to more advanced forms of reporting?

FINDINGS

Three different cases of news automation will be presented next within a discursive framework for the relationship between journalistic work and technology developed by Matthew Powers (2012).

Journalists have during the last four decades assessed the impact of new technology in their work in three distinct ways: (1) as exemplars of continuity; (2) as threats to be subordinated; and (3) as possibilities for journalistic reinvention (Powers, 2012). Journalists tend to discuss "technologically specific" forms of work, a term which according to Powers refers to "work rooted in the affordances of technical capacities that also make claims about the journalistic nature of such work" (Powers, 2012, p. 25). This study gives practical examples of all three discourses in advanced computational journalism. These cases are not comparable in between and do not contain any grand narratives of the changing nature of journalism, but they exemplify the three different traits that have been laid out.

Case 1: Associated Press
Software generated news is used to ensure the continuity of news operations by increasing efficiency with automation of certain editorial tasks, such as laborious, monotonous, and error-prone routine duties.

In New York in early 2014, editors at the Associated Press, a news agency formed in the spring of 1846, started testing simple algorithms from software provider Automated Insights in Durham,

North Carolina, that generated earnings reports on listed companies. The underlying data was provided by Zacks Financial Services in Chicago, Illinois. The AP has struggled with downsizing for a number of years while customers have expected reporters to maintain the same volume of wire news. One of the most demanding editorial tasks has been coverage of earnings reports. The "earnings season" occurs four times a year when listed companies publish their quarterly earnings reports. Financial journalists at AP used to be tied up for the whole day probing through the numbers. AP's standards editor Tom Kent, a veteran journalist at the news agency, came up with an ethical checklist (Kent, 2015) for software-generated news. His final piece of advice in the list is:

> The best protection as you move further into robot newswriting is a constant focus on testing, and on making sure editors understand how the software really works. Plus recognition that many things are still best done by humans.

With intensive testing, AP went to lengths to make sure the language in the automatically generated texts was rich enough, adhered to the established AP style guide, and exhibited domain-specif-

ic features, so that, for instance, financial news did not look like sports news. Thus, the newsroom was looking to create something more than templates with blanks to fill, like Mad Libs, the phrasal template word game. A news manager at AP said in the interview:

> It's a much more creative act than that. It's based on analysis, and then a real writing process [...] and that's very important, because I think our subscribers and readers would go crazy if every story read exactly the same way. [...] They're supposed to be stories, they're not supposed to be reformatted data.

With every article produced automatically by AP comes a link that explains how the piece has been generated and where the data comes from.

Internal evaluation of automation shows, according to the interviews, that earnings stories contain fewer spelling or calculation errors than man-made articles, that the output has gone from around 300 stories per earnings season to more than 3,700, and that reporters are more satisfied than before. The processing time from when the press release arrives to when the news item reaches the customer is anywhere between one and ten minutes. A news manager said:

> I was shocked at how easy it turned out to be. [...] I've just been really surprised [...] that things weren't harder than they've been or that they weren't impossible.

A human reporter reads a limited number of earnings reports carefully to analyse implications for the American economy since the performance of certain companies such as John Deere or Caterpillar is an indicator of general trends. A few large and complex companies such as Citigroup and Salomon Brothers are only handled by human reporters.

AP has been able to save financial resources and move money into the travel and entertainment budget. A news manager said:

> I definitely see a difference in terms of the flow of work and the lift in the newsroom, the lack of drudgery. One day I walked into the area, I'd been in a meeting, and one of the desk editors was standing up talking to someone at 9:40 in the morning and I thought to myself, "During earnings season before, you never ever would have seen someone up, having a conversation, unless there was a problem."

This case can arguably be analysed as an example of how technology becomes a part of continuity.

Case 2: Local Labs

Automation of journalism's routine tasks can also be framed as a threat materialising in loss of journalist jobs. However, in this case automation enables small town papers to survive in print and the creation of new ones.

Local Labs is a small company in Chicago, Illinois founded in 2006 that operates small town newspapers on its own and provides local editions and other services for metro newspapers in suburbs. The company was formerly known as Journatic and changed its name after a controversy around fake bylines and other ethical issues (Romenesko, 2014; Tarkov, 2012). Local Labs unifies and centralises editorial processes with its combination of a newsroom management system with partly automated content creation and proactive gathering of events information. The operational logic is that an American town of 100,000 inhabitants has roughly 20,000 news sources that are people and organisations arranging events. By approaching these sources with automated emails, which are triggered by a rationale based on seasonal celebrations or other regularly occurring events, newsrooms are

less dependent on the memory capacity of individual reporters. A manager at the company says:

> This technological solution is about fewer journalists going and asking and the machine facilitating it so that the knowledge and the memory in the business are in the institution, not in the individual journalist.

[ ] @

Local Labs is minimizing the employment of journalists and has moved certain functions like newspaper design to low-cost countries like the Philippines. The business logic is that reactive journalism costs money that can be saved by building proactive editorial solutions, for instance anticipatory processes or developing software that fills the newspaper pages automatically without any human intervention. One representative of the company explains: "If you're reactive every week, then it costs more money."

Local Labs uses different combinations of journalists and machines to produce content – sometimes human only, sometimes machine only, sometimes hybrid production forms. The representative says there is no way to take man out of the equation, but it is possible to make man more efficient. However, the idea that machines will become smart enough to replace journalists is in his view out of the question. According to the manager,

> The idea that we're going to take unstructured data someday, it's never going to happen ever. It's a pipe dream. Algorithms only work on structured data. That's it. They only work on structured inputs. That's true of any computer. You can't take unstructured inputs and structure them on the fly.

Local Labs is using automation and new ways of organising work, including outsourcing to low-cost countries, to offer solutions for publishers at a very attractive price but reducing the numbers of journalists needed to run a news operation.

Case 3: ProPublica
Automation can also create new forms of employment as part of the reinvention of journalism.

ProPublica, a non-profit news organisation in New York founded in 2007, decided in 2012 to experiment with software-generated news together with Narrative Science. A news application, Opportunity Gap, was created in a few weeks and produced more than 52,000 stories about the quality of American schools. The evaluation received mixed views. One of the data journalists involved explained why this probably should be a one-off project:

> When we write a story with numbers, we painfully pick the right words to say each number and report about it. I guess maybe for some things like a sports story or weather story that are exactly the same every time, that that could save resources. But our experience was […] going through and checking all that stuff was a nightmare. It was an interesting experience, but not one I would want to repeat.

The assistant managing editor Scott Klein, who leads the News App team at ProPublica, explained in a blog post (Klein, 2013) how journalists worked together with engineers to tune algorithms and find the right style, a slow and cumbersome process.

From the other side, people involved in designing the automated news application experienced the same concerns that were presented in the literature review – namely, journalists are not able to explain in detail the work they are doing. A computer scientist noted:

We thought that the journalists could tell us what is the structure of a baseball game recap story. They couldn't. They can do it, but they can't express it explicitly. It's not part of the culture of journalism to be reflective on the structure of the genre. [...] The reason we were surprised is that I think that for other literary pursuits, it is. If you write poetry or short stories, the kinds of poems there can be and the kinds of short stories there are and the way, in which the logic of a story is propelled forward, I think that's a lot of what somebody who goes to study how to write. [...] One of the things we realized is that journalism isn't that way because 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. The writing is not even important at some level.

The real world as expressed in digitally stored data must be standardised and unified in a formal manner, otherwise the algorithms will not be able to process and generate news. From one data scientist's point of view it is understandable that there is so much unusable data, for instance in public records:

Eighty per cent of any data project is standardisation, normalisation, and validation. It's cleaning the data up and making it vaguely useful. Everyone's so sloppy about data but ... it doesn't really matter if there are gaps because nobody was using it for anything.

"Cleaning" data means detecting and removing errors and inconsistencies from data as well as acting upon missing information or misspellings in order to improve the quality of data (Rahm & Do, 2000). The integration of data from different sources also requires extensive manual work. At ProPublica a team of ten programmers work with the data, of which only one person has a degree in journalism, but the whole newsroom seems to be very data literate based on the interviews. ProPublica also sells cleaned data sets. Even if this only produces small revenues it shows how new forms of journalism are emerging.

CHALLENGES FOR JOURNALISTS

In the first research question it was asked how managers, journalists, and programmers working with news innovations anticipate the future of software-generated news and what do they think about the role of journalists. The interviews show that computer scientists are quite happy to work with journalists because they know they might get to work with data which is hard to access. Quite a few journalists have become experts on filing Freedom of Information Act (FOIA) requests and through their practices open up access to new data on a regular basis. Journalists working with public data seem to assume that open access data generally is of limited interest, that the information that would generate stories is stacked away. An experienced data journalist agreed:

For most of the stories we do, we still go and get data from an agency. It's not stuff that they make available online.

Bots and drones also collect data, but these operations need to be designed and monitored by someone and the data used in a meaningful way. In this sense the constraints on software-generated news are grounded both in the limited capacity to create algorithms that understand unstructured and messy data as well as in the lack of access to data that is reasonably structured to allow less ad-

vanced algorithms to create automated content, and this should not be seen as a minor hindrance.

Another constraint is statistical and the way journalists, or rather their editors, treat the concept of probability. An editor would certainly not accept a news item with an 85% or 95% certainty of being true, while dealing with uncertain issues such as climate change inherently means accepting that there are qualified estimations we might not be able to prove 100%.

There are also concerns that automation of news will block young people from entry level jobs where they are able to learn how to do journalism and acquire tacit knowledge. A news manager expressed his concerns:

> You have to learn your trade by doing when the stakes are low. You learn how to read an earnings statement. You learn how to talk to people, you learn how to talk to police and you learn how to watch a baseball game from the point of view of a reporter and not a fan. I worry about what it might do.

Research on computational journalism has emerged out of the nexus between computer science and journalism and primarily been concerned with "building things" (Anderson, 2013), which

is only natural considering the new skills that are needed. A cultural obstacle exists in journalism practice that can be illustrated by a quote from one academic involved in both journalism and computer science.

> They're [journalists] very skeptical to the point [that] in a technology community they would be considered negative. They're always down to, "That won't work. Why do you think that will work?? Why do you think that's good enough?" [Computer scientists say:] I don't know but shut up and let's build it.

It seems like journalists are used to being in charge of editorial processes. Overarching social and cultural concerns need to be part of the reorientation towards advanced computational journalism. One practical example: documentation of work progress is mainly absent in journalism methods while in computer science it is crucial to have accurate metadata that, for instance, explains how data has been collected, as well as detailed descriptions on how software code has been written. That is an integral part of computer science, so that the next person will be able to understand and use the same items of data or code.

FUTURE OF ALGORITHMS AND JOURNALISM

The second research question – can the promise that automation will free journalists from repetitive routine tasks to more qualitative forms of journalism really be trusted – is partly possible to answer based on the expert interviews. At AP this seemed definitely to be the case, while it is less clear if this is true with Local Labs. Therefore I turn to the research literature for additional guidance and I will focus on some positive expectations of computational journalism, namely that human journalists will need to work smarter with the help of better algorithms, developing hybrid forms of human-machine interaction much in the way that chess players have improved their ratings thanks to smart chess programs (Thompson, 2010) or pathologists use computer aided diagnostics (CAD) to improve efficiency, accuracy, and consistency (Hipp et al., 2011). The impact on journalism does not stop at software-generated news based on structured data from financial reports or sports results. Self-writing software will be able to rewrite its functionality on the move, "whether through genetic algorithms, viral coding structures or merely connective or artificial intelligence like encoding behaviour or expert systems" (Berry, 2011). Computer scientists

have for the last couple of decades been working on evolutionary algorithms whose purpose is to solve multi-objective optimisation problems that involve several conflicting objectives (Zhou et al., 2011). The evolutionary algorithm learns how to solve problems with a reasonable level of competency even when the conditions change. These algorithms, though still in an early stage, may provide developers with solutions to automate more advanced forms of journalism, a system that adapts its behaviour to meet its goal in a range of environments (Domingos, 2015; Fogel, 2006). However, there have been many false starts of human-level artificial intelligence where systems have failed to perform robustly on real-world tasks (Bostrom, 2014).

I suppose that algorithms will force journalists to think harder at defining their core human capabilities, such as developing emotional and social intelligence, curiosity, authenticity, humility, empathy, and the ability to become better listeners, collaborators, and learners. Journalists tend to define their occupation by the tasks that are fulfilled rather than what skills and knowledge a person needs to possess to fulfil them (Van Dalen, 2012). Subject matter experts will also continue to exist outside the realm of what we can expect from ro-

bots (Hoyle Brown, Roehrig, & Malhotra, 2015).

However, journalists need some basic training in computational thinking (CT) which refers to a way of "solving problems, designing systems, and understanding human behaviour that draws on concepts fundamental to computer science" (Wing, 2006, p. 33). CT is the integration of logical, algorithmic, scientific, and innovative dimensions of human cognition, pointing towards openness to new ideas, revolutionising all kinds of occupations, including journalism (Gynnild, 2014; Wing, 2011). In essence, these are new domain skills of journalism for mastering data, the product of automation and digitisation. The CEO of Narrative Science, Stuart Frankel, has said that the few human writers who work for the company have to become "meta-journalists" who design the templates, frames and angles into which the algorithm inserts the data (Gore, 2015). Because of the explosion of rich process-level data, journalists need different analytical domain skills to develop and work with artificial intelligence and machine learning (Hoyle Brown et al., 2015).

If people practicing journalism are capable of thinking along those lines they should also be able to expand the traditionally rather narrow self-perception of professional opportunities. After all,

their capabilities will cover expertise that is central to all adults: key information-processing skills in literacy (reading skills), numeracy (mathematical skills), and skills in problem-solving in technology-rich environments (OECD, 2013).

CONCLUSIONS: AUTOMATION AND NEWS WORK

Here are some provisional conclusions from this study. Journalism logic on its lowest level can easily be built into algorithms and thus form a simple automaton model for journalism, while more complex forms of journalism are harder to turn into news applications due to the high degree of uncertainty around the application of rules.

Data is a core issue. News algorithms cannot function without structured data and the way data is chosen, evaluated, and "cleaned" – standardised, normalised and validated, some would say manipulated – is, or at least should be, a process subject to journalistic decisions. There are many questions regarding the accessibility of data, where it comes from, how it is produced, and for what purpose. The futurist dream that all public data will be open source is contradicted by the fact that, for instance,

public bodies such as municipalities don't have the resources or the incentives to digitize printed records. The access to data becomes an economic, not a democratic, decision. However, there are plenty of private data producers who, for a number of reasons, some of them less altruistic, are more than willing to let media companies use their data. In this case a critical journalistic approach is necessary.

News algorithms function as self-contained processes and "black boxes" but they are socially constructed (Bijker, Hughes, & Pinch, 1987), thus they are subject to negotiation. That social context is a cause of concern to be discussed. Journalists should therefore be able to add value to the process of building news applications by providing crucial meaning-making on what is important and significant – distilling insights. Holding people in power to account is somewhat less trivial than helping users find the best cup of coffee in the neighbourhood.

Journalists certainly need to deal with intraorganisational tension and work closely with computer scientists or programmers in the process of creating algorithms for different purposes, but should also learn how to investigate algorithm decision systems critically.

The relationship with the audience is a special concern (Napoli, 2003). Journalists have with the arrival of citizen journalism already been moving away from a transactional relationship with the audience to an interactional exchange, but rather as one-off and ad hoc. Journalists now have, thanks to user data, much better analytical tools for understanding user engagement with certain types of content and services in multiple and complex ways. Editorial decisions are increasingly being based on large sets of user data and supported by automated processes. This means that media organisations will need to fundamentally redefine what media audiences mean to them (Napoli, 2012).

This study within an institutional framework of media production has primarily shown how the work of journalists is empowered and supplemented, but also replaced by smart machines. The three cases of news automation studied showed 1) increased efficiency and job satisfaction with automation of monotonous and error-prone routine tasks (Associated Press), 2) automation of journalism routine tasks resulting in losses of journalist jobs (Local Labs), and 3) new forms of work that require computational thinking (ProPublica). One main finding is that rule-based instructions func-

tion in a finite sense in software-generated news, mainly due to the lack of structural data, but might not be applicable to all forms of journalism. Even though the chief scientist of Narrative Science, Kris Hammond, claims that computers will generate stories that win the Pulitzer Prize in a few years (Levy, 2012) the results of present operations with automated news shows the limitations: real artificial intelligence is still a work in progress, including evolutionary algorithms that are being developed. Journalists tend to work with certain rules that are historically grounded in the media production logic, but that also inform ethical decisions. However, these are inconsistent and flexible, which reflects the need for creative and intuitive judgments in dealing with uncertainty and social interaction. In this sense, there is no algorithm for advanced journalism.

Instead, the combinations of computer processing power, programming, and human collective creativity have produced impressive results in journalism judging by recent projects based on large data sets, such as Reading the Riots, Cracking the Codes, Luxleaks, WikiLeaks' Iraq War Logs, Docs for Dollars, the Panama Papers, and others.

There are certainly ethical, moral, and operational considerations that come with the advent of software-generated news, since journalistic aspects are only part of the picture. Publishers, advertisers, data producers, governments, and users all have their own agendas in the political economy of algorithm systems. To this could be added that artificial intelligence tends to concentrate power in fewer hands, as already has been seen with Google, Facebook, and Twitter.

Algorithms can also be manipulated to an extent that is not possible with human beings, not even with the most aggressive attempts of coercion. These are some of the fields where journalism and media researchers need to think hard about the algorithm issue, especially the capabilities and impact of journalists.

REFERENCES

Acemoglu, D., & Robinson, J. A. (2012). *Why nations fail: The origins of power, prosperity and poverty.* New York: Crown Publishers.

Akst, D. (2013). Automation anxiety. *The Wilson Quarterly, Summer* Retrieved from http://archive.wilsonquarterly.com/sites/default/files/articles/AutomationAnxiety.pdf

Anderson, C. (2011). Notes towards an analysis of computational journalism. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2009292

Anderson, C. (2013). Towards a sociology of computational and algorithmic journalism. *New Media & Society, 15*(7), 1005-1021.

Bakhshi, H., Frey, C. B., & Osbourne, M. (2015). *Creativity vs. robots: The creative economy and the future of employment.* London: Nesta.

Berry, D. M. (2011). *Philosophy of software: Code and mediation in the digital age.* London: Palgrave Macmillan.

Bijker, W., Hughes, T., & Pinch, T. (Eds.). (1987). *The social construction of technological systems: New directions in the sociology and history of technology.* Cambridge MA: MIT Press.

Bogner, A., Littig, B., & Menz, W. (2009). Introduction: Expert interviews—An introduction to a new methodological debate. In A. Bogner, B. Littig & W. Menz (Eds.), *Interviewing experts* (pp. 1-13). Basingstoke: Palgrave Macmillan.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies.* Oxford University Press.

Clerwall, C. (2014). Enter the robot journalist. *Journalism Practice, 8*(5), 519-531. doi:10.1080/17512786.2014.883116

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.

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

Cohen, S., Li, C., Yang, J., & Yu, C. (2011). Computational journalism: A call to arms to database researchers. Paper presented at the *CIDR, 2011,* 148-151.

Cook, T. E. (1998). *Governing with the news: The news media as a political institution.* Chicago: University of Chicago Press.

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.

Edge, A. (2014). Ophan: Key metrics informing editorial at the Guardian. Retrieved from https://www.journalism.co.uk/news/how-ophan-offers-bespoke-data-to-inform-content-at-the-guardian/s2/a563349/

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 79*(10), S70-S81.

Ess, C. M. (2014). Editor's introduction: Innovations in the newsroom – and beyond. *Journal of Media Innovations, 1*(2), 1-9.

Fairfield, J., & Shtein, H. (2014). Big data, big problems: Emerging issues in the ethics of data science and journalism. *Journal of Mass Media Ethics, 29*(1), 38-51. doi:10.1080/08900523.2014.863126

Flew, T., Spurgeon, C., Daniel, A., & Swift, A. (2012). The promise of computational journalism. *Journalism Practice, 6*(2), 157-171. doi:10.1080/17512786.2011.616655

Fogel, D. B. (2006). *Evolutionary computation: Toward a new philosophy of machine intelligence.* Hoboken, New Jersey: John Wiley & Sons.

Frey, C. B., & Osborne, M. (2015). *Technology at work: The future of innovation and employment.* Oxford: Oxford Martin School and Citigroup.

Frey, C. B., & Osborne, M. A. (2013). *The future of employment: How susceptible are jobs to computerisation?* (No. 103). Oxford: OMS Working Papers.

Garrison, B. (2001a). Computer-assisted reporting near complete adoption. *Newspaper Research Journal, 22*(1), 65-79.

Garrison, B. (2001b). Diffusion of online information technologies in newspaper newsrooms. *Journalism, 2*(2), 221-239.

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski & K. Foot (Eds.), *Media technologies, essays on communication, materiality and society* (pp. 167-194). Cambridge, MA: MIT Press.

Gore, A. (2015). *The future: Six drivers of global change.* New York: Penguin Random House.

Grint, K., & Woolgar, S. (1997). *The machine at work: Technology, work and organization.* Cambridge: Polity Press.

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., & Turner, F. (2009). Accountability through algorithm: Developing the field of computational journalism. Paper presented at the *Summwe Workshop, 27-41.* Retrieved from http://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf

Hanitzsch, T. et al. (2011). Mapping journalism cultures across nations: A comparative study of 18 countries. *Journalism Studies, 12*(3), 273-293.

Hermida, A., & Thurman, N. (2008). A clash of cultures. *Journalism Practice, 2*(3), 343-356. doi:10.1080/17512780802054538

Hipp, J. et al. (2011). Computer aided diagnostic tools aim to empower rather than replace pathologists: Lessons learned from computational chess. *Journal of Pathology Informatics, 2,* 25-3539.82050. Epub 2011 Jun 14. doi:10.4103/2153-3539.82050

Howard, A. B. (2014). *The art and science of data-driven journalism.* New York: Tow Center for Digital Journalism, Columbia University.

Hoyle Brown, R., Roehrig, P., & Malhotra, V. (2015). *The robot and I: How new digital technologies are making smart people and businesses smarter by automating rote work.* Teaneck, NJ: Cognizant.

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

Kent, T. (2015). An ethical checklist for robot journalism. Retrieved from https://medium.com/@tjrkent/an-ethical-checklist-for-robot-journalism-1f41dcbd7be2

Kidder, R. M. (1995). *How good people make tough choices*. New York: Morrow.

Kim, J., Lee, K., Kim, Y., Kuppuswamy, N. S., & Jo, J. (2007). Ubiquitous robot: A new paradigm for integrated services. Paper presented at the *2007 IEEE International Conference on Robotics and Automation,* 2853-2858.

Klein, S. (2013). How to edit 52,000 stories at once. Retrieved from http://www.propublica.org/nerds/item/how-to-edit-52000-stories-at-once

Kvale, S. (2005). The dominance of dialogical interview research. *Impuls, 1*, 5-13.

Latar, N. L. (2014). Robot journalists: 'Quakebot' is just the beginning. Retrieved from http://knowledge.wharton.upenn.edu/article/will-robot-journalists-replace-humanl-ones/

Levy, S. (2012, 24 April). Can an algorithm write a better news story than a human reporter? *Wired*.

Lewis, S. C., & Westlund, O. (2014). Big data and journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism, 3*(3), 447-466. doi:10.1080/21670811.2014.976418

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

Maier, S. R. (2000). Digital diffusion in newsrooms: The uneven advance of computer-assisted reporting. *Newspaper Research Journal, 21*(2), 95-110.

March, J. G., & Olsen, J. P. (1989). *Rediscovering institutions: The organizational basis of politics*. New York: Free Press.

Matsumoto, R., Nakayama, H., Harada, T., & Kuniyoshi, Y. (2007). Journalist robot: Robot system making news articles from real world. Paper presented at the *2007 IEEE International Conference on Robotics and Automation,* 1234-1241.

McLuhan, M. (2011). *The Gutenberg galaxy: The making of typographic man* University of Toronto Press.

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *The American Journal of Sociology, 83*(2), 340-363.

Meyer, P. (1973). *Precision journalism: A reporter's introduction to social science methods*. Bloomington: Indiana University Press.

Napoli, P. M. (2012). Audience evolution and the future of audience research. *International Journal on Media Management, 14*(2), 79-97.

Napoli, P. M. (2014a). Automated media: An institutional theory perspective on algorithmic media production and consumption. *Communication Theory, 24*(3), 340-360.

Napoli, P. M. (2014b). On automation in media industries: Integrating algorithmic media production into media industries scholarship. *Media Industries, 1*(1)

Napoli, P. M. (2003). *Audience economics: Media institutions and the audience marketplace*. New York: Columbia University Press.

Nunez, M. (2016). Want to know what Facebook really thinks of journalists? Here's what happened when it hired some. Retrieved from http://gizmodo.com/want-to-know-what-facebook-really-thinks-of-journalists-1773916117

OECD. (2013). *Skills outlook 2013: First results from the survey of adult skills*. Paris: OECD Publishing.

Örnebring, H. (2016). *Newsworkers: A comparative European perspective*. London: Bloomsbury Academic.

Örnebring, H. (2010). Technology and journalism-as-labour: Historical perspectives. *Journalism, 11*(1), 57-74. doi:10.1177/1464884909350644

Parasie, S., & Dagiral, E. (2012). Data-driven journalism and the public good: "Computer-assisted-reporters" and "programmer-journalists" in chicago. *New Media & Society, 15*(6), 853-871. doi:10.1177/1461444812463345

Pavlik, J. (2000). The impact of technology on journalism. *Journalism Studies, 1*(2), 229-237. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=ufh&AN=3887851&site=ehost-live&scope=site

Polanyi, M. (1967). *The tacit dimension.* London: Routledge & Kegan Paul.

Powers, M. (2012). "In forms that are familiar and yet-to-be invented" american journalism and the discourse of technologically specific work. *Journal of Communication Inquiry, 36*(1), 24-43.

Quinn, A. (2007). Moral virtues for journalists. *Journal of Mass Media Ethics, 22*(2-3), 168-186.

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. *IEEE Data Eng.Bull., 23*(4), 3-13.

Romenesko, J. (2014). Hyperlocal content creator journatic is renamed locallabs. Retrieved from http://jimromenesko.com/2014/04/14/hyperlocal-content-creator-journatic-is-renamed-locallabs/#respond

Saltzis, K., & Dickinson, R. (2008). Inside the changing newsroom: Journalists' responses to media convergence. *AP, 60*(3), 216-228. doi:10.1108/00012530810879097

Singer, J. B. (2004). Strange bedfellows? the diffusion of convergence in four news organizations. *Journalism Studies, 5*(1), 3-18.

Sparrow, B. H. (1999). *Uncertain guardians: The news media as a political institution.* Baltimore: Johns Hopkins University Press.

Stavelin, E. (2014). *Computational journalism. when journalism meets programming* (PhD).

Stone, H. S. (1971). *Introduction to computer organization and data structures.* New York: McGraw-Hill, Inc.

Tarkov, A. (2012). Journatic worker takes 'This American Life' inside outsourced journalism. Retrieved from www.poynter.org/news/mediawire/179555/journatic-staffer-takes-this-american-life-inside-outsourced-journalism/

Thompson, C. (2010, March 22). Clive Thompson on the cyborg advantage. *Wired.*

Thorén, C. (2014). *Print or perish? A study of inertia in a regional newspaper industry* (PhD).

Tunstall, J. (1975). *Journalists at work: Specialist correspondents: Their news organizations, news sources, and competitor-colleagues* (Repr. ed.). Beverly Hills, Calif: Sage.

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.

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33-35.

Wing, J. M. (2011). Computational thinking. Paper presented at the *OurCSWorkshop, 3.*

Young, M. L., & Hermida, A. (2014). From mr. and mrs. outlier to central tendencies. *Digital Journalism, 3*(3), 381-397. doi:10.1080/21670811.2014.976409

Zhou, A., Qu, B., Li, H., Zhao, S., Suganthan, P. N., & Zhang, Q. (2011). Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm and Evolutionary Computation, 1*(1), 32-49.