sciendo

Nordic Journal of
MEDIA STUDIES

Unboxing news automation

Exploring imagined affordances of automation in news journalism

Stefanie Sirén-Heikel[I], Leo Leppänen[II], Carl-Gustav Lindén[III] & Asta Bäck[IV]

[I] Faculty of Social Sciences, University of Helsinki, Finland, stefanie.siren-heikel@helsinki.fi
[II] Department of Computer Science, University of Helsinki, Finland, leo.leppanen@helsinki.fi
[III] Swedish School of Social Science, University of Helsinki, Finland, and Södertörn University, Sweden, carl-gustav.linden@helsinki.fi
[IV] VTT Technical Research Centre of Finland Ltd, Finland, asta.back@vtt.fi

Abstract

News automation is an emerging field within journalism, with the potential to transform newswork. Increasing access to data, combined with developing technology, will allow further inquiries into automated journalism. Producing news text using NLG (natural language generation) is currently largely undertaken in specific, predictable news domains, such as sports or finance. This interdisciplinary study investigates how elite media representatives from Finland, Europe and the US imagine the affordances of this emerging technology for their organization. Our analysis shows how the affordances of news automation are imagined as providing efficiency, increasing output and aiding in reallocating resources to pursue quality journalism. The affordances are, however, constrained by such factors as access to structured data, the quality of automation and a lack of relevant skills. In its current form, automated text generation is seen as providing only limited benefits to news organizations that are already imagining further possibilities of automation.

Keywords: automated journalism, news automation, newswork, NLG, imagined affordances

Introduction

News automation has emerged in recent years as a technology with the potential to provide a new means of production for the news media industry. Recent developments in natural language generation (NLG), increased access to structured data and disruptions affecting the revenue model of the media landscape have created an environment in which automation is being considered for predictable news

Sirén-Heikel, S., Leppänen, L., Lindén, C-G. & Bäck, A. (2019). Unboxing news automation: Exploring imagined affordances of automation in news journalism. *Nordic Journal of Media Studies*, 1: 47-66. doi:10.2478/njms-2019-0004.

stories. Several international and national news organizations have begun producing news articles programmatically based on standardized data (LeCompte, 2015; Wu et al., 2018). News automation (NA), also known as automated journalism, converts structured data into text based on a set of rules. It is limited to specialized areas: the current systems are employed in well-understood domains, such as finance, sports and election reporting. Nonetheless, there is an interest in conceptualizing, developing and deploying automated journalism in news production.

This article aims to capture a specific moment of technological emergence by analysing the way in which media representatives talk about the imagined affordances of news automation in relation to the production of news in their organization. Accordingly, we combine two sets of empirical accounts: 1) interviews with elite representatives from a diverse set of European and American media organizations and 2) a set of in-depth interviews with media representatives conducted as part of a Finnish research and development project studying news automation. The material is from 2015 to 2018 and showcases organizations at different stages of considering, developing and deploying the technology.

We aim to add to the burgeoning research on automation in newsrooms[1] by focusing our study on the affordances of the technology. Our interest lies in identifying how media representatives envision and make sense of this developing technology and its relevance to their organization. Considering the emergent nature of the technology, it is the relationship between the material constraints of the technology and the social practices of news that shapes the definition of this new form of production. To study this in-between phase, we apply the concept of affordances, more specifically imagined affordances (Nagy & Neff, 2015), as a framework for our analysis. The concept offers scholars a middle ground between technological determinism and social constructivism, a Latourian "third way" (Hutchby, 2001) that focuses on the framework that media representatives envision around news automation.

Literature review

Digital disruption and the media landscape

Technological developments have brought considerable changes to the global media landscape during the last few decades. Historically, technological breakthroughs have often affected the media industry by altering the production – and our conceptualization – of communication and news (Chadwick, 2013; Pavlik, 2000). At the same time, the arrival of new technology has often caused anxiety and uncertainty (Ryfe, 2012). Digitalization has transformed the way in which content is produced and consumed, reorganizing the workforce in newsrooms with a heightened focus on computational skills (Lewis & Usher, 2016). Chadwick (2013) pointed to how the emergence of a new technology tends to be viewed from a linear perspective, with the new inevitably replacing the old technology, when examples from history show an interactive process wherein each shapes the

other (25-26). Technology does not create change in a vacuum: the organizational culture, audience habits, ownership structures, socio-economic factors and infrastructure play their part.

Commercial news media organizations have seen a major part of their advertising revenue shift to a handful of global tech companies, a decrease in subscription revenue and dependence on digital platforms for distribution (Gill et al., 2013; Küng, 2017). In the aftermath of the financial crisis that began in 2008, both commercial and public service newsrooms suffered cutbacks in their workforces (Doctor, 2015; Ryfe, 2012). The restructuring of newsrooms to support cross-media production has added to the conflict between journalistic norms and practices and organizational pressures, requiring more to be achieved with limited resources (Hofstetter & Schoenhagen, 2017). This disruption has reshaped audiences into quantifiable, trackable consumers, with audience metrics influencing editorial decisions (Arenberg & Lowrey, 2018; Petre, 2015). The introduction of software metrics is "enhancing the effectiveness of news content", infusing newsrooms with profit-oriented norms and values (Belair-Gagnon & Holton, 2018: 13). At the same time, analyses have shown that the conversion of users into loyal customers is dependent on a mix of variant high-quality journalistic content (Jungkvist, 2018; Shishkin, 2017). It is in this milieu that the potential for using automated processes for the production of news stories in such domains as finance and sports is becoming of interest to media organizations.

Automated text generation for news journalism

Automated journalism is usually defined as the use of automated computer systems to produce textual news articles based on structured data. Some authors have applied the term strictly to systems that run without human intervention outside the original set-up of the system (Graefe, 2016), while others have included systems that require some human intervention (Carlson, 2015). Either way, automated journalism in practice currently refers to the use of NLG methods in the domain of news.[2] While the first academic research into the use of NLG to produce factual reports is decades old, particularly in the domain of weather reports (Glahn, 1970, 1979; Goldberg et al., 1994), commercial actors' application of NLG to news is a more recent phenomenon (Dörr, 2016).

NLG systems are either *rule based* or *machine learning based*. Systems based on rules generate text based on a series of man-made rules, whereas systems based on machine learning learn rules by observing significant amounts of learning material. The rule-based system is limited by its requirement for specific rules for all (reasonable) articles in a specific domain, which are time consuming to produce. The machine learning-based system is constrained by its need for a massive amount of learning material, and it can only learn rules that are present in the material. Employing a machine learning system is largely unfeasible in complex domains with a limited span of human-written articles. Thus far, the

commercial application of news automation has largely been limited to a few well-understood domains with little internal variability, for example weather (Lindén, 2017), earthquake reports (Oremus, 2014), sport and property news (United Robots, n.d.), and corporate earnings calls (Coldford, 2014). What these domains have in common is that they all report strictly on an event or quantifiable facts. These domains focus mainly on telling us what happened or is happening. The limitation of only answering the "what", rather than the "why", is due to the inability of computer systems to analyse events against contextual life-world knowledge.

A key issue with respect to automation is its dependency on high-quality standardized data (Leppänen et al., 2017; Wu et al., 2018), a situation that has been eased somewhat by digitalization, which has facilitated the datafication of society (Mayer-Schönberger & Cukier, 2013). The idea of datafication is to turn aspects of daily interaction and relationships into quantifiable data traces and metadata: transportation, social media behaviour, purchases, searches, geolocation and other forms of activity in which applications or software collect data on human or non-human behaviour. Existing data are increasingly being made open, offered by service providers, by organizations or by governments for third-party use (Dietrich et al., 2012). Datafication supposedly allows more "sophisticated" analysis of large data sets by breaking down information into data, thereby enabling a better understanding and better services (Mai, 2016: 193). In the current era, in which digitalization is evolving into datafication and more and more data are being made open and available, developing and deploying automated systems can become technologically relevant to news.

Affordances of automation

Studies of news automation in newsrooms have often framed the change in newswork "vis-à-vis the machines" (Wu et al., 2018: 1) as a "technological drama" between the technology and the journalistic practices (Carlson, 2015) and as comparisons between computer-written and human-written texts (Graefe et al., 2016). These approaches to studying humans and technology tend (un)intentionally to shift towards determinism, whether social or technological (Grint & Woolgar, 1997). As automated systems are based on performing specific tasks according to specific rules with a specific outcome in mind, as defined by humans, it is relevant to acknowledge this relationship (Broussard, 2018). This viewpoint has risen to the forefront in recent years in works by scholars dissecting big data and algorithmic decision-making (see, e.g., boyd & Crawford, 2012; Eubanks, 2018; Pasquale, 2015). In conjunction with this, the determinist neutrality of big data has been described as a "myth" (Couldry, 2017) and seemingly objective algorithm models as "opinions embedded in mathematics" (O'Neil, 2016: 21).

Using the concept of imagined affordances when examining news automation enables a focus on the multidirectional relationship between humans and

technology, avoiding a polarizing unidirectional dichotomy of either technological determinism or social constructivism (Faraj & Azad, 2012). The concept of affordances is well established in technology studies and has gained popularity in media and communication research.[3] Nonetheless, the definitions of the concept have been highly fluid and polyvalent. Originally introduced in the late 1970s in ecological psychology by J.J. Gibson (1979), and further developed in cognitive design studies in the 1980s by Donald A. Norman (2013), the concept is generally applied to describing what a material artefact allows, its actionabilities and its perceived usability. For Gibson, an affordance is invariant and exists regardless of whether anyone sees a way to use it, it is "neither an objective property nor a subjective property; or it is both if you like" (1979: 129), it either exists or it does not, it is independent of prior knowledge and it is relational. After a few iterations, Norman similarly understood affordances as a "relationship between the properties of an object and the capabilities of the agent", and this joint interaction determines the affordance (2013: 11). However, for Norman, an affordance can be misunderstood, difficult to enact and contextual. This view also encompasses designing hindrances and constraints.[4]

Despite its theoretical fluidity, the concept can "provide a useful tool for user-centered analyses of technologies" (Gaver, 1991: 79) and is applicable to the study of the "symbiotic relationship between people and technology" (Ahuja et al., 2018: 2204). Blewett and Hugo argued for a Latourian acceptance of affordances in which the focus is on the "in-between", framing our understanding "within an actant <-> actant relationship" (Blewett & Hugo, 2016: 65). We argue that the same inseparability, or "entanglement" (Orlikowski, 2007), is present when conceptualizing, designing and deploying NA in media organizations. Traditionally, affordances have been conceived of in relation to such material features as surfaces, substances or items affording actions. In our digitalized world, affordances are increasingly intangible, accessible through digital artefacts. These digital artefacts offer what Leonardi (2010) defined as a "practical instantiation", meaning, for example, software embodying materiality through its significance for action (para. 5). As an example, an automated system affords a newsroom certain actions by combining data with a set of algorithms – but the data have to be collected through material interactions and the software has to be designed.

To design something, it has to be imagined. From the perspective of our explorative study on how media representatives envision an emerging technology, the concept of imagined affordances is useful (Nagy & Neff, 2015). By looking at the in-between of technologies and humans, both the "material and the perceptual" (Nagy & Neff, 2015: 2), we attempt to bridge automation with newswork. The imagined affordances are formed through the perceptions of what the technology affords and not necessarily what it allows. As the possibilities and limitations of NA are uncertain and developing, these visions can be seen as "recognizing the unexpected, situated and emergent" in the possibilities of engagement with

the technology (Faraj & Azad, 2012: 252). Nagy and Neff's concept allows for a focus on the "adaptation in practice and in interaction" (2015: 5) of a technology, as both the technology and its use are mouldable, not inherent.

However, our imagination is shaped by bits and pieces of life-world experience, social and material, which shape our ability to reassemble expectations. Alternatively, as Gibson put it in regard to perception, "it is the very features of the object that your perceptual system has already picked up that constitute your ability to visualize it" (Gibson, 1979: 257). From this standpoint, affordances are seen as sociotechnical, occurring within the relationship between the social principles and needs (i.e. the journalistic norms and practices) and the technological framework (i.e. the data and computer processing). "Imagination connotes perception, not just rationality" (Nagy & Neff, 2015: 5), which is relevant in the context of understanding how the participants in our study framed their views on news automation.

Methodology

To investigate how media organizations imagine the affordances of news automation, we explored the way in which media representatives talk about the technology. For our interviews, we focused on elite representatives in positions in which they have the power to conceptualize, develop and deploy automation in their organization. This article combines interviews with European and American actors, in the vanguard of employing new technology, with in-depth interviews with Finnish media representatives, resulting in a rich, diverse body of material. When analysing the material, our aim was to find out how media managers imagine the affordances of news automation and, consequently, which factors frame the way in which the affordances are imagined?

Interview design

The interviews with European and American media representatives (n=14) were collected during the period 2015-2018 as material for studying digital disruption in the news media industry, and they functioned as a starting point for the Finnish research project on news automation. Snowball sampling was used as a method for identifying relevant participants (Noy, 2008). The interviews varied in length, ranging from approximately fifteen minutes to one hour, and in the number of participants, ranging from one to three. Some were conducted over Skype, whereas others were recorded during conferences or other types of meetings.

The in-depth interviews with Finnish media representatives (n=12) were conducted in 2017-2018 as a part of the research project "Immersive Automation", an academic R&D project studying news automation. All of the media organizations that participated in the project were part of a consortium connected to the research project and had either implemented or considered implementing

some form of NA. Representatives from the Finnish public service company YLE were added to the list of interviewees to give a more characteristic crosscut of the Finnish media landscape. The participants were chosen primarily by identifying persons in the organizations in positions of influence over production processes, with a focus on editors or managers. Most participants were chosen by the researchers based on insights into the organizations, while others were added through recommendations from the organizations. The project-related interviews were all face-to-face interviews, and their duration varied between 45 minutes and 90 minutes. In one of the interviews, more than one participant from the organization was present.

As this is an emerging field, a semi-structured format was applied to the interviews to allow for open and revealing responses in relation to NA. The interviews related to the research project had a tighter question framework than the ones conducted outside the project, which were looser and more oriented towards the specifics of the organization that the participants represented. The interviews with the European and American participants included such questions as: How do the newsrooms discuss what journalists should do after you have added these methods? What could be the model for news media in terms of artificial intelligence? The interviews with the project participants included such questions as: How do you see automation changing the quality of information in five years? Will automation increase or decrease competition?

Analysis design

The interviews were digitally recorded, transcribed and analysed using Atlas.ti for sets in the material relevant to the research questions, which were further analysed actively to seek out themes related to the topic of imagined affordances (Braun & Clarke, 2006). The interviews were initially coded row by row, with the coding emerging from the material, reflecting our theoretical approach. The reply structures tended to be circular and explorative, favouring recoding with conceptual codes and descriptive sub-codes for greater analytical cohesion. Instead of looking at keywords, the material was analysed for meaning and underlying contexts and structured into themes identified to respond to our research interests (Braun & Clarke, 2006). The themes reported in this article are the ones that reached saturation and validity in relation to our theoretical framework. The thematic analysis was conducted by the main author, who is fluent in the languages in which the interviews were conducted: Finnish, Swedish and English. The selected quotations have been translated into English by the main author.

Table 1. Study participants

ID	*Position*	*Media Type*	*Type of automation at date of interview*	*Date*
A	Senior Digital Manager	Regional media company	None	2017
B	Senior Digital Manager	National media company	Tools for story leads	2017
C	Deputy Digital Manager	Public service broadcaster	Domain-specific, template-based NLG	2018
D	Senior Digital Manager	National media company	Tools for story leads, automated b-to-b services	2017
E	Senior Digital Manager	Regional media company	In discussion with service provider for NLG texts	2018
F	Editor	Regional media company	In discussion with service provider for NLG texts	2018
G	Senior Manager	National media company	Domain-specific, template-based NLG	2017
H	Managing Editor	Regional newspaper	None	2017
I	Deputy Digital Manager	National media company	Tools for story leads	2017
J	Developer	Public service broadcaster	Domain-specific, template-based NLG	2017
K	Managing Editor	Public service broadcaster	Domain-specific, template-based NLG	2017
L	Managing Editor	National media company	Tools for story leads	2017
M	Editor	International news agency	Domain-specific, template-based NLG	2015
N	Editor	National newspaper	Tools for story leads	2015
O	Developer	Independent regional news site	Tools for story leads	2015
P	Developer	Independent regional news site	Tools for story leads	2015
Q	Editor	Independent regional news site	Tools for story leads	2015
R	Senior Manager	National online newspaper	None	2018
S	Senior Manager	National independent news agency	Domain-specific, template-based NLG	2018
T	Senior Digital Manager	National business newspaper	Domain-specific, template-based NLG	2018
U	Senior Digital Manager	Regional media company	Domain-specific, template-based NLG	2016
V	Managing Editor	National investigative news media	Domain-specific, template-based NLG	2015
W	Senior Manager	Multinational news media company	Domain-specific NLP, sentiment analysis etc.	2015
X	Editor	International news agency	Domain-specific, template-based NLG	2015
Y	Editor	International news agency	Domain-specific, template-based NLG	2015
Z	Senior Digital Manager	Regional media company	Domain-specific, template-based NLG	2016

Comment: As to avoid identification, positions have been standardised via general descriptions, and country of media organisation omitted.

Findings

In the previous sections, we outlined the aim of our study, the theoretical framework and the methodology. In this section, we present the findings from our analysis to explore our research questions. Media organizations operate in a disruptive environment in which they need to evolve and adapt to gain and retain audiences and secure their place in the landscape. These issues are reflected in the imagined affordances, which are organized thematically here according to their different aspects; whether they are easy or difficult to achieve, occur as part of a sequential continuum or occur simultaneously with other affordances, or misinterpretations, leading us to the factors framing the way in which news automation is imagined.

Degrees: From easy to difficult

Departing from Gibson's binary view of affordances, McGrenere and Ho (2000) introduced the notion of *degrees* in connection with affordances. They referred to ease of use and clarity of information in design, but the idea is applicable when considering the range of usefulness of NA. When considering the features of the software systems, some parts can be easy, whereas others can prove to be difficult. The imagined usability might not match the usefulness. The participants expressed a fairly level-headed view of what NA allows for in its current state. The technology can increase the number of stories but only for specific topics with suitable data. As pointed out by participant D, there is no "Jesus technique" that could "conjure" a story on any subject from any type of source.

An NLG system based on recurring data with set patterns can write volumes of sports stories, but it cannot explain why a referee accepted an offside goal. "There is sort of a limited story angle set that you will want to write, which is, 'So-and-so won the football match. So-and-so lost the football match. It was a draw'", commented participant S. One of the main affordances of NLG, the speed of reporting, is diminished due to the need for human intervention when the system fails to understand the context, as participant K explained:

> At least according to current understanding, a machine helps in getting greater volumes, but when you ought to […] go deeper into the subject to interpret and draw conclusions, at least at this point you need a human.

In its current form, automation is not seen as replacing human journalists, but, as suggested by participant W, it will probably lead to journalists being assigned "to more of the higher value-added activities". One constraint of automation is the difficult, time-consuming task of creating a fluent, variable language. Another is its predictable nature, as exemplified by participant S, wherein the systems make decisions based on rules such as "this is a story because that number is bigger than that one". There is a sense that letting technology drive the newsroom strategy is not particularly beneficial or imaginative. Participant B noted that "data in itself

does not bring anything new to the table", unless it is connected with contextualization. For this reason, participants envision using automation as a tool for journalists and not only as a direct product for audiences. This leads to visions of the further affordances of news automation.

Sequential: A chain of affordances

One of the aims of the efforts to automate "is to make life easier for journalists in newsrooms. We don't want to give them a job, we want to give them something which liberates them", explained participant S. News automation is imagined as improving and augmenting the existing reporting, releasing journalists from repetitive tasks. In this sense, the imagined affordances are *sequential* (Gaver, 1991), whereby "acting on a perceptible affordance leads to information indicating new affordances" (Gaver, 1991: 82). To approximate Gibson, a large rock provides not only the actionability of climbing but also improved vision, which again might reveal something that standing on the ground would not. Participant X stated: "I'm in favour of anything that removes drudgery from our journalists and enables them to do real journalistic work"; as participant J said, automation could offer a "tremendous advantage" in supporting newsrooms.

The datafication of society is understood to influence newswork, but the sheer volume of data poses difficulties. Employing automation is imagined to aid core newswork by sifting for leads or producing reports from data sets. The automated, increased output is viewed as making possible story angles and topics that previously would have been unattainable with the existing resources. Automation is seen as allowing new forms of content for audiences. Participant U explained:

> We saw a demand around sports … weather and traffic. There is a huge interest around this. What we see now is a demand around housing. In our analysis, that's what people are willing to pay most for.

The relationship with audiences is further considered through the credibility that automation can provide, strengthening brand reputation and trustworthiness. In an era driven by opinion, platformonomics and fake news, participant F suggested that automated stories represent "facts […] and figures, not someone's manipulated interpretation". Participant K commented that automation can help journalists write stories that are increasingly "comprehensive, multifaceted and weighing of all sides". News automation can alleviate errors in stories, as long as the data are of a high quality. According to participant S, it can perhaps aid journalists to "recalibrate a little bit the impressions they have of their own community", taking them from "anecdotal into empiric data".

Other sequential affordances reflect the changing realities of media production, with visions of producing audio and even video from text, as discussed by participant B: "The digitalization of voice […] interests us very much. And that is also one field of automation; how do we then convert the content we have written

[…] into speech?" There are visions of automation that provide for more versatile, and particularly personalized, services. Participant B described how aggregation and automated translations of content have potential as new niche businesses when targeted directly at customers.

Nested: A complex system

Personalization is an imagined affordance that displays *nested* features: data collection on the user, which again aids further in automated story generation, targeting and the conceptualization of new niches. Participant U explained that "we want insights into what the user wants", which further facilitate the embedding of emotional wording (especially in sports), targeted at inducing reactions in the user. "It's about solving a problem for the user", continued participant U, and "automated text is one way to do it". Whereas sequential affordances refer to situations in which information indicates new affordances, *nested* affordances are "grouped in space" (Gaver, 1991: 82). Nested affordances exist simultaneously, as in the case of personalized news stories, in which text generation, personalization and ad targeting are interconnected. Personalization has been around for as long as news itself (Chadwick, 2013); however, platform affordances and metrics have heightened its importance. Participant J discussed how this key feature of streaming platforms can be enabled in news as well through automation:

> That's where I think that we are headed as well in regard to journalism … [You] get your own topics and even your own articles and your own snippets and you can embed audio and video […] if that's what you prefer.

However, personalization requires diversified content, meaning several angles of the same story. When newsrooms attempt to meet the demands for quality and story depth, releasing the workforce to create volumes of content becomes a discrepancy. Participant K explained how this contradiction will push them towards using news automation: "Personalization will forcibly take us there […] automation and robotics will make it easier for us when we can have the robot write some of the content". Personalization is not only imagined as benefiting the organization from within; it is nested with affordances for users struggling with a "flood of information": "with automation, this information can find me … [instead] of me having to find this information", said participant A. Personalization is connected to an increase in advertising, as the volume of content fine-grains the targeting. On the other hand, participant D commented on how the ideal of a segment-of-one approach is problematic, as "not all [ad space] buyers are ready for this [idea of] 'let's personify my message for each and every one' to a person on a personal level".

To encourage advertisers to personalize ads and interest audiences, the technology has to combine several factors, such as access to trustworthy data, quality output and high-speed distribution. Participant S explained the nested structure of

the system as "a big build around with an NLG component sitting in the middle of our process". Another bundled system exists within that particular component, in which the functionality of the language is as important as the algorithm that identifies the outliers that decide the stories.

False: Misleading misinformation

As a continuation from degrees of affordances, the respondents were conflicted about how they imagined the usefulness of news automation. Borrowing from Gaver (1991), such affordances can be described as *false* in the sense that they are perceived but perhaps with limited or misinterpreted information. Gibson (1979) used the term *misinformation* for a situation in which an actor picks up information that is partial or an illusion. Norman (2013) would probably have called them *misleading signifiers*, as he separated affordances from perceptible signifiers that signal actionabilities. In any case, the issue at hand is a mismatch between the perceived abilities of a technology and what it affords, which several of the participants recognized and discussed. Participant B, understanding the technology, commented that it is necessary to keep in mind the limitations of automation in terms of understanding variability: "Let's say that I'm not interested in sports 364 days of the year, but one day per year I am, because it happens to be [about] my sister". The current systems are not able to perceive such sudden changes.

Framing the usefulness of news automation is its ability to give something to the organization, whether as input or as output, without homogenizing the content, "threatening the diversity" or diminishing the distinctiveness of the news product. Participant R, from an organization that at the time of the interview was not using NA, discussed how the product defines the perceived affordances of automation; a weekly is different from a newspaper or a business site: "You expect something different. I don't see that yet being done by automated news." Some went as far as rejecting the affordances of NLG as being too simplified and unimaginative. The current affordances do not "raise the pulse"; in fact, automated text can become "impossibly boring" if media organizations become gripped by a "trance" of automation and focus only on efficiency instead of building services that fascinate, a warning heeded by participant G. Furthermore, participant L suggested several other, more interesting, forms of automation than just news bulletins "about ice hockey":

> I'm fascinated by the idea connected to different ways of weeding, combing, refining, distributing the material and how this material could, for example, be transformed into different forms: texts, audio and images, graphics, these kinds of things.

Participant C concluded that "this text form and its production, will be left […] as one transitional period, but the developments will certainly not stop there". Even as the affordances imagined via NA are relevant, text generation is only

seen as one path to automation, in particular as there is already an abundance of text. "There is no end in itself for robot texts", said participant U, explaining that NLG is useful only if it has value for users. Participant G commented that it is "nice to have" software that can produce text from data but that currently the articles generated rarely "result in cheers" from the audiences. For context, you currently need a journalist, argued participant Q:

> It's one thing to say who wins a basketball game; it's another thing to say this donation indicates a pattern of influence that connects with years […] of history about how the city does... I mean, that's at least right now, not a thing a computer can do. Maybe theoretically they could, but you need a ton of data.

To avoid falling victim to false promises, deciding what to use automation for is a key question. "Does it promote anything in our lives, really?", asked participant L. Shaving a few minutes off writing might not be such a great affordance after all, particularly in journalism, in which speed is good but "accuracy is more important". Overall, news automation is viewed as a continuation of digitalization and processes that have been present in newsrooms for decades, just "fancier". The respondents shared an understanding that overt technicism can cause an urge to invest in shiny new things and use data just because they are there, without actually providing anything meaningful.

The ambiguous nature of news automation

The accounts of the respondents revealed an undercurrent of uncertainty regarding how news automation should be imagined. Automation adds a new layer of disruption to the envisioning of strategies for future products and services, which has already been a struggle for organizations engaging with digitalization (Küng, 2017). One participant wondered whether his or her organization had enough "visionary capacity" to decide which affordances are worthy of investment. Estimating the unknown implies that some hidden affordances "must be inferred from other evidence" as much as possible (Gaver, 1991: 80). The respondents discussed the benefits and the drawbacks from a broad perspective: how it affects their organization, the media industry and civil society. Even though the affordances are imagined as being tightly coupled with improving the existing reporting, participant J noted how that idea hinges on newsroom funding:

> If the drastic downsizing continues and is compensated for with […] bulk automation, [then] there is also a risk that the news coverage will be impoverished.

If the journalistic reporting is not up to par, the willingness to pay in the face of cheaper, mass-produced, automated news may diminish. One imagined scenario is new media businesses based on automation, producing content cheaply, without the requirement of running a newsroom. Participant A argued that, if automated systems allow "millions of articles a minute" to be produced, the consequences

can be hard to predict. Some expressed concern that issues that are already affecting the credibility of the media might be amplified via automation: information warfare, misinformation and the siloing of users.

Another consideration concerns the building blocks of NLG, the data. One question is that of exclusive rights to data collected by private entities, which again might further limit the possibilities for smaller organizations with limited resources and skills. Data can be expensive, fragmented and difficult to obtain; the sheer volume, velocity and variety require the capacity to process it. "Our job is to scrutinize those in power", stated participant U, but that job becomes difficult if data are not made open or structured. Media organizations do not necessarily have the skills or the resources to handle large sets of data securely, verify them and ensure that they do not violate privacy. Participant B pondered how the media should deal with sensitive data, such as the pulse rates of football players, or how they should ensure a chain of accountability if they cannot trace and verify the data themselves. With systems that have the capacity to produce thousands of articles in a heartbeat, an "error [can] metastasize itself through the whole thing", said participant X. An unsolved question concerns where the responsibility lies in such cases: with the data provider, the system provider or the publisher?

If data are one issue, another is the competence to use them in a fruitful manner: "It requires that we know how to pose relevant questions for relevant data", commented participant H. Due to unclear practices of liability, organizations must grasp such issues as the explainability and the transparency of the software, which currently "isn't going to happen" according to participant G. Whereas newsrooms understand journalism, analysis and context, most do not have proficiency in understanding data and algorithmic decision-making tools such as NLG. For this reason, they need outside system providers that can produce relevant software: systems that fit the needs of the newsrooms, are easy to use and are adapted to the norms and values of journalism and the identity of the organization. Even so, automation can require large investments and hands-on work with service providers to fine-tune the outcome. Even then, participant J noted that rollouts of new technologies in rigid media organizations pose their own difficulties, "even if it in theory would be possible to revolutionize a line of business completely".

These issues reflect the main factors influencing how news automation is imagined: access to data, skills and resources to use and understand the data and a system that produces output that is of good enough quality. Who comprises the audience is equally important: a news agency and a daily local would not benefit from similar automation. Participant L wondered on how well short news snippets made by NLG really can entice audiences in such a competitive field as news media. On the other hand, the efficiency promised by news automation can be attractive for struggling media organizations, which, according to participant A, might "be forced to implement automated solutions".

Predictable, routine stories that make journalists "feel like robots" are most suitable for automation. Participant S commented on the ease of "interrogating

a sheet of data" for stories, exemplifying how news automation has connections with data journalism: "this wasn't about automation at all at the start; it was about trying to get journalists to use open data as a source because there might be stories there which were easy to get at." In the end, it is about financial realities: at the moment, most organizations have both to exploit their current audiences and explore new strategies to gain and retain customers. Participant V summed up the situation as follows: "I think [automation is] good to have, if it helps news organizations stay in business and if it helps them pay for more complex things". Since news automation and NLG are emerging, evolving technologies, they make for "exciting times" when newswork and news automation can interact, as long as the "the human touch" of journalism remains at the core.

Discussion

In this exploratory study, we have attempted to highlight the imagined affordances of news automation and the underlying factors framing the way in which participants envision the technology. News automation is an emerging technology, deriving from digitalization through datafication, and is currently appearing in newsrooms. Newswork is a profession in which several fields of interest intertwine: affordances are imagined through the norms and practices connected to news journalism, and the financial and business realities of the organisations. The affordances imagined reflect Nagy and Neff's (2015) conceptualization of affordances as features that can be "present for only one individual or a group of individuals but not for others" (Nagy & Neff, 2015: 3). News automation is imagined by the participants as variant and multi-faceted, combining concrete low-level affordances with abstract high-level affordances (Bucher & Helmond, 2018), mirroring the value-based expectations regarding the actions that automation can afford (c.f. Nagy & Neff, 2015). Conversely, this can be seen as a way of normalizing a new technology by situating it within the existing frames of newswork, "squelching the potential for fundamental change" (Lewis & Usher, 2016: 547). What the participants imagine and expect indicates their power, or lack thereof, to interpret the affordances (Nagy & Neff, 2015). In this sense, we situate the concept of affordances in the middle ground between Gibson and Norman: as relational yet contextual, intertwining the social and the technological.

Following certain definitions of artificial intelligence (AI), newsrooms are already filled with AI tools affording actionabilities and performing them autonomously: spam filters, word processors and search functions. Many future advancements will likely fall into the category of "impossible before, obvious afterwards". Automation can be seen as part of the same continuum that replaced the pen with first a typewriter and then a computer. Advances with the current systems are likely to be relatively gradual. NLG based on machine learning, especially if it is multilingual, can result in more significant and dramatic changes. If automation becomes established in newsrooms, it may move journalism further

towards editorial expertise and contextualization, answering the "why". Its current ability to provide only a "what" is holding back the adoption of automated journalism in a more significant fashion.

One issue concerns access to dependable data, whether the data are collected in a structured manner or affordable for media organizations to use. Furthermore, despite recent work towards multi-domain and multi-language NLG systems in academia (Leppänen et al., 2017), commercial NLG is still seen as being prohibitively expensive to implement outside a very few domains (Lindén, 2017), and it is limited by a lack of resources and competence. In summary, the affordances of news automation have to provide media organizations with clear benefits to be worthwhile. As this is an exploratory study, future research is needed to study the longitudinal impact of news automation – and other forms of automation in the newsroom – on newswork. Does automation in effect lead to the affordances envisioned? If so, will organizations reallocate resources to deeper, qualitative reporting, influencing the product, audiences and business models? Our study warrants further fieldwork to observe how these imagined affordances are adopted and enacted in practice and how they affect newswork.

Limitations of the study

Despite our attempt to apply the concept of imagined affordances as a means of avoiding unfruitful dualism, our study contains traces of "residual technicism" (Grint & Woolgar, 1997: 37). The research design set out to analyse the imagined aspects of a technology as a labour-saving device in newsrooms, reflecting a predominant discourse of automation and its prowess, whereas the relationship between the social and the technical in this context is more complex (Elish & boyd, 2018). The changes affecting the way in which news is produced, consumed and financed are not occurring as part of an isolated process; rather, they are part of a larger phenomenon altering society as a whole. In addition, relying on a few representatives of organizations does not provide insights into the organizations' "collective imagination" of automation (Ryfe, 2009: 677). Our approach of locating relevant representatives through recommendations and snowballing poses the risk of inadvertently omitting social groups that can be relevant to the development of news automation (Klein & Kleinman, 2002).

Conclusion

Our analysis of the way in which media organizations imagine the affordances of news automation show the technology to be ambiguous. It can provide for sequential effects, such as reallocating resources to investigative journalism, improved reporting and new product niches. It can alleviate a need for volume with respect to personalization, but at the same time it requires bundled solutions, such as distribution and access to structured data. The prospects of news

automation are dependent on resources, skills, organizational identity, norms and values. Imagining the affordances of automation for news is challenging, as the discourse surrounding artificial intelligence and big data is often mystified and embellished. Audience-facing NLG in particular is viewed as limited in its scope, mainly solving repetitive tasks and providing new services at best. The affordances are imagined as posing certain threats to the existing media business models and to the credibility of journalism.

The usability of news automation appears to trump its usefulness. Currently, automation in newsrooms for text generation is at an early stage, but, even so, human–computer co-creation and interaction will remain the locus for news automation for the foreseeable future. The questions are whether media representatives have the capacity to imagine the affordances of news automation, whether the information for the affordances is available for perception and whether the affordances in fact are there to be perceived in the first place. However, the imagined affordances-as-practice can also be different from those that are currently available, indicating that the realities of newsrooms – and users – could be taken into account better when conceptualizing and developing systems for news automation. Separating the social from the technical, and vice versa, is an unproductive approach to examining the potential of news automation. Attaching too much expectation to the technology risks a situation in which the "affordances suggest different actions than those for which the object is designed", resulting in errors, possible misalignment of the actual uses and declining interest (Gaver, 1991: 5). Similarly, disregarding the social obscures the ambitions, powers and desires behind the technological systems, which, in the current media landscape, can have a profound effect on media producers and users.

Funding

This work has been co-funded by Business Finland (previously TEKES, the Finnish Funding Agency for Technology and Innovation), Svenska kulturfonden, the Media Industry Research Foundation of Finland and a consortium of Finnish media companies, the University of Helsinki and VTT Technical Research Centre of Finland Ltd as part of the research and development project "Immersive Automation", an interdisciplinary project conducted jointly by the Swedish School of Social Sciences at the University of Helsinki, the Department of Computer Science at the University of Helsinki and the VTT Technical Research Centre of Finland Ltd between January 2017 and May 2018.

Acknowledgements

The interviews with Finnish media representatives were designed and conducted jointly by Professor Emeritus Tom Moring and Principal Scientist Asta Bäck, with the assistance of doctoral student Stefanie Sirén-Heikel. The interviews with European and American representatives were designed and conducted by Adjunct Professor Carl-Gustav Lindén. A work-in-progress version of this article was presented at the EMMA annual conference in June 2018.

Notes

1. See, for example, the special issue of *Digital Journalism*, "Algorithms, automation, and news: Capabilities, cases, and consequences" (forthcoming).
2. For a technical survey of the state of the art in NLG, see Gatt and Krahmer (2018).
3. Among others, Costa (2018) introduced the concept of affordances-in-practice when studying how context collapse on Facebook is mitigated in Turkey. Tenenboim-Weinblatt and Neiger (2018) studied how the affordances of news technologies influence journalistic storytelling practices.
4. For further elaboration of the concept and its iterations, see Blewett and Hugo (2016), Bucher and Helmond (2018) and McGrenere and Ho (2000).

References

Ahuja, M., Patel, P. & Suh, A. (2018). The influence of social media on collective action in the context of digital activism: An affordance approach. In *Proceedings of the 51st Hawaii International Conference on System Sciences* (pp. 2203-2212). NY, USA: Hawaii International

Arenberg, T. & Lowrey, W. (2018). The impact of web metrics on community news decisions: A resource dependence perspective. *Journalism & Mass Communication Quarterly*, 1-19. doi: https://doi.org/10.1177/1077699018801318.

Belair-Gagnon, V. & Holton, A. E. (2018). Boundary work, interloper media, and analytics in newsrooms: An analysis of the roles of web analytics companies in news production. *Digital Journalism*, 6(4): 492-508. doi: https://doi.org/10.1080/21670811.2018.1445001.

Blewett, C. & Hugo, W. (2016). Actant affordances: A brief history of affordance theory and a Latourian extension for education technology research. *Critical Studies in Teaching and Learning*, 4(1): 55-73.

boyd, D. & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5): 662-679. doi: https://doi.org/10.1080/1369118X.2012.678878.

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2): 77-101.

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. Cambridge, MA: MIT Press.

Bucher, T. & Helmond, A. (2018). The affordances of social media platforms. In J. Burgess, A. Marwick & T. Poell (eds.), *The SAGE handbook of social media* (pp. 233-253). Sage Publications.

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, 3(3): 416-431. doi: https://doi.org/10.1080/21670811.2014.976412.

Chadwick, A. (2013). *The hybrid media system: Politics and power*. Oxford: Oxford University Press.

Coldford, P. (2014, June 30). A leap forward in quarterly earnings stories. *The Definite Source* [blog]. Retrieved from https://blog.ap.org/announcements/a-leap-forward-in-quarterly-earnings-stories. [accessed 2014, November 15].

Costa, E. (2018). Affordances-in-practice: An ethnographic critique of social media logic and context collapse. *New Media & Society,* https://doi.org/10.1177/1461444818756290.

Couldry, N. (2017). The myth of big data. In M. T. Schäfer & K. van Es (eds.), *The datafied society* (pp. 235-239). Amsterdam: Amsterdam University Press.

Dietrich, D., Gray, J., McNamara, T., Poikola, A., Pollock, R., Tait, J. & Zijlstra, T. (2012). The open data handbook: An open knowledge foundation project. Retrieved from http://opendatahandbook.org/ [accessed 2018, February 21].

Doctor, K. (2015, July 18). Newsonomics: The halving of America's daily newsrooms. *Nieman Lab* [online]. Retrieved from http://www.niemanlab.org/2015/07/newsonomics-the-halving-of-americas-daily-newsrooms/ [accessed 2018, December 15].

Dörr, K. N. (2016). Mapping the field of algorithmic journalism. *Digital Journalism*, 4(6): 700-722. https://doi.org/10.1080/21670811.2015.1096748.

Elish, M. C. & boyd, D. (2018). Situating methods in the magic of big data and AI. *Communication Monographs*, 85(1): 57-80. doi: https://doi.org/10.1080/03637751.2017.1375130.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press.

Faraj, S. & Azad, B. (2012). The materiality of technology: An affordance perspective. In P. M. Leonardi, B. A. Nardi & J. Kallinikos (eds.), *Materiality and organizing: Social interaction in a technological world* (pp. 237-258). Oxford: Oxford University Press.

Gatt, A. & Krahmer, E. (2018). Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 61: 65-170. https://doi.org/10.1613/jair.5477.

Gaver, W. W. (1991, April). Technology affordances. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 79-84). ACM.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin Company.

Gill, P., Erramilli, V., Chaintreau, A., Krishnamurthy, B., Papagiannaki, K. & Rodriguez, P. (2013). Follow the money: Understanding economics of online aggregation and advertising. In *Proceedings of the 2013 Conference on Internet Measurement* (pp. 141-148). http://dx.doi.org/10.1145/2504730.2504768.

Glahn, H. R. (1970). Computer-produced worded forecasts. *Bulletin of the American Meteorological Society*, 51(12): 1126-1131.

Glahn, H. R. (1979). Computer worded forecasts. *Bulletin of the American Meteorological Society*, 60(1): 4-11.

Goldberg, E., Driedger, N. & Kittredge, R. I. (1994). Using natural-language processing to produce weather forecasts. *IEEE Intelligent Systems*, 2: 45-53.

Graefe, A. (2016, January 6). Guide to automated journalism. *Tow Reports* [online]. Retrieved from https://www.cjr.org/tow_center_reports/guide_to_automated_journalism.php [accessed 2017, March 7].

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

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

Hofstetter, B. & Schoenhagen, P. (2017). When Creative Potentials are Being Undermined by Commercial Imperatives. Digital Journalism, 5(1): 44–60.doi: http://doi.org/10.1080/21670811.2016.1155966

Hutchby, I. (2001). Technologies, texts and affordances. *Sociology*, 35(2): 441-456. https://doi.org/10.1177/S0038038501000219.

Jungkvist, K. (2018). Digital subscriptions – A challenge for survival. Paper presented at the seminar *Vem betalar för journalistiken?* [Who will pay for journalism?]. 2018, December 14, at the Swedish School of Social Sciences, University of Helsinki.

Klein, H. K. & Kleinman, D. L. (2002). The social construction of technology: Structural considerations. *Science, Technology, & Human Values*, 27(1): 28-52.

Küng, L. (2017). *Strategic management in the media: Theory to practice*. London: Sage Publications.

LeCompte, C. (2015, September 1). Automation in the newsroom: How algorithms are helping reporters expand coverage, engage audiences, and respond to breaking news. *Nieman Reports* [online]. Retrieved from https://niemanreports.org/articles/automation-in-the-newsroom/ [accessed 2017, February 22].

Leonardi, P. M. (2010). Digital materiality? How artifacts without matter, matter. *First Monday* [online], 15(6). doi: https://doi.org/10.5210/fm.v15i6.3036.

Leppänen, L., Munezero, M., Granroth-Wilding, M. & Toivonen, H. (2017). Data-driven news generation for automated journalism. In *Proceedings of the 10th International Conference on Natural Language Generation* (pp. 188-197). doi: http://doi.org/10.18653/v1/W17-35.

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

Lindén, C. G. (2017). Decades of automation in the newsroom: Why are there still so many jobs in journalism? *Digital Journalism*, 5(2): 123-140. https://doi.org/10.1080/21670811.2016.1160791.

Mai, J. E. (2016). Big data privacy: The datafication of personal information. *The Information Society*, 32(3): 192-199. https://doi.org/10.1080/01972243.2016.1153010.

McGrenere, J. & Ho, W. (2000, May). Affordances: Clarifying and evolving a concept. In *Proceedings of Graphics Interface* (pp. 179-186).

Nagy, P. & Neff, G. (2015). Imagined affordance: Reconstructing a keyword for communication theory. *Social Media+Society*, 1(2): 1-9. doi: https://doi.org/10.1177/2056305115603385.

Mayer-Schönberger, V. & Cukier, K. (2013). *Big data*. London: John Murray.

Norman. D. A. (2013). *The design of everyday things*. Cambridge, MA: MIT Press.

Noy, C. (2008). Sampling knowledge: The hermeneutics of snowball sampling in qualitative research. *International Journal of Social Research Methodology*, 11(4): 327-344.

O'Neil, C. (2016). *Weapons of math destruction*. New York: Crown Publishing Group.

Oremus, W. (2014, March 17). The first news report on the LA earthquake was written by a robot. *Slate* [online]. Retrieved from http://www.slate.com/blogs/future_tense/2014/03/17/quakebot_los_angeles_times_robot_journalist_writes_article_on_la_earthquake.html. [accessed 2014, November 21].

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies*, 28(9): 1435-1448.

Pasquale, F. (2015). *The black box society*. Boston: Harvard University Press.

Pavlik, J. (2000). The impact of technology on journalism. *Journalism Studies*, 1(2): 229-237. https://doi.org/10.1080/14616700050028226.

Petre, C. (2015, May 7). The traffic factories: Metrics at chartbeat, gawker media, and the New York Times. Retrieved from https://www.cjr.org/tow_center_reports/the_traffic_factories_metrics_at_chartbeat_gawker_media_and_the_new_york_times.php/ [accessed 2017, August 18].

Ryfe, D. M. (2009). Structure, agency, and change in an American newsroom. *Journalism*, 10(5): 665-683.

Ryfe, D. M. (2012). *Can journalism survive? An inside look at American newsrooms*. Cambridge, UK: Polity Press.

Shishkin, D. (2017, July 3). Five lessons I learned while digitally changing BBC World Service. *LinkedIn* [blog]. Retrieved from https://www.linkedin.com/pulse/five-lessons-i-learned-while-digitally-changing-bbc-world-shishkin/ [accessed 2018, November 15].

Tenenboim-Weinblatt, K. & Neiger, M. (2018). Temporal affordances in the news. *Journalism*, 19(1): 37-55. https://doi.org/10.1177/1464884916689152.

United Robots. (n.d.). United Robots: Products. Retrieved from http://unitedrobots.ai/produkter-1/ [accessed 2019, 11 January].

Wu, S., Tandoc Jr, E. C. & Salmon, C. T. (2018). Journalism reconfigured: Assessing human–machine relations and the autonomous power of automation in news production. *Journalism Studies*, 1-18. doi: https://doi.org/10.1080/1461670X.2018.1521299.