AUTHOR

**Lauri Haapanen**

# MEDIA COUNCILS AND SELF-REGULATION IN THE EMERGING ERA OF NEWS AUTOMATION

JULKISEN SANAN
NEUVOSTO

# Media councils and self-regulation in the emerging era of news automation

▶ LAURI HAAPANEN

**THIS REPORT INTRODUCES** the present state of affairs in news automation and discusses what ethical considerations it raises in the work of media councils. Based on a European-wide research project, the three key takeaways are as follows:

1) Large media players use automatically updated counters and infographics, but the automated generation of news text is still mainly experimental in nature. We need to keep our eyes open to the possible need for self-regulatory guidance on news automation, though there are as yet no urgent issues.

2) We must not underestimate the technical progress already made in the field. Issues regarding data, agency and transparency should not be overlooked when discussing news automation today.

3) Media councils should critically revisit their complaints procedures so that audiences have a genuine opportunity to bring up their grievances with automation. Most importantly, the councils should act proactively with regard to the changing media landscape.

The key is that if it is not the media councils that take the lead on this, it is going to be someone else. And whoever it is – whether national legislators, the EU or platform companies – they might jeopardise the freedom of the press. This means that external control and guidance might force journalistic media to make decisions about content and publishing on non-journalistic grounds.

**KEYWORDS:** Algorithms, codes of ethic, journalism, media councils, news automation, personalization, self-regulation, transparency



KUVA: JOUNI JYLLINMAA

**ABOUT THE AUTHOR**

Lauri Haapanen (PhD) is University Lecturer in Journalism Studies at the University of Jyväskylä, Finland. He has worked in the projects *Immersive Automation* (2017–2018), which created a roadmap and a demonstration of a future news ecosystem based on news automation; and *Targeted Journalism* (2020–2022), which examines the ways, goals and the future of personalization. He has also been a member of the Council of Mass Media in Finland (2017–2019), participating in the working group that prepared an advisory statement on marking news automation and personalization.

# Table of Content

# Introduction:
# The future that is already here

Everyone in the media industry, and other people too, have already heard of *news automation* and seen some examples of it, whether they know it or not. The term refers to algorithmic processes that convert data into a user-friendly form with limited human intervention. Speaking from a perspective that may still be a little premature, **news automation can – when working well – speed up production, increase the breadth of coverage, enhance accuracy and enable new types of personalization**.

Here, *speeding up* means that an algorithm-driven system is able to generate news stories in a fraction of a second. *Increasing the breadth* indicates that we do not have to be happy with a few general news stories but, for example in the context of elections, software can now generate a text about the results of each polling station separately. When it comes to *accuracy*, the computer does not get tired or try to finish quickly because it is time to go home. *Personalization* means that different versions of a news story are made according to the interests of each user – for example, telling the reader how the party they personally support did in the election.

**News automation refers to algorithmic processes that convert numerical data into a user-friendly form with limited human intervention.**

Rather than such *full-blooded* news automation, i.e. software-driven generation of news text, **news automation so far takes place mainly in the form of various infographics and up-to-date counters, the main emphasis of which is on numerical information instead of long stretches of text**. The coronavirus pandemic and the U.S. presidential election have made these applications familiar to everyone this year.[1]

Personalization is often referred to not just as a part of news automation but as a separate process. It is practical because its applications are primarily about the user's previous online behaviour affecting "read more" recommendations or, less commonly, the order in which news headlines are displayed on the front page of a web page. It is like an automated search function into the archive of that media that reads the user's thoughts. **To sum up, the personalization that has taken place so far affects mainly the logic in which media operators offer their articles to their users, and only rarely the content of individual articles.**

---

1 An example of this is the New York Times' monitoring of the coronavirus (**https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html**) and elections (these counters are already closed).

# How has self-regulation responded to news automation?

In media self-regulation, automation has so far received little direct attention. **News automation or the use of algorithms more generally is not mentioned in ethical codes.**[2] An online survey that I conducted with the Chairs of 35 European media councils, of whom 21 responded,[3] indicated that **media councils are waiting and watching**, the most common response being that "news automation is more or less non-existent in our media market, and also news personalization is very rare". However, it is worth noting that media councils do not have a comprehensive picture: media operators implement algorithm-driven solutions as they feel like it, with no obligation to report to councils or the authorities. (**See a selection of answers in BOX 1**.)

About one in three respondents said that although the question is not yet acute, they have already discussed **the challenges news automation and personalization might pose and the possible need for guidance** (See **BOX 2** on next page). However, the Finnish Council for Mass Media is so far the only council that has taken more visible steps to take a stand on news automation. In late 2019 it issued a *Statement on marking news automation and personalization*,[4] which supplements the current *Guidelines for Journalists*.[5] I will return to this later.

In what follows, I will consider, first, whether the present state of the art of news automation calls for self-regulatory action. To spoil your excitement a bit, my answer – the first takeaway of this report – is,

---

## BOX 1

▶ There is almost no news automation in the Hungarian market, and personalization is also very rare. Facebook's algorithm, of course, has a huge effect on news consumption, though.

▶ Personalization is fairly widely used in many kinds of media outlets. The media in Finland is also very keen to develop and apply new personalization tools. News automation is used less and only in the media outlets with the largest financial resources.

▶ Unfortunately I cannot answer the question, there are no statistics.

▶ News automation: none
(to the best of my knowledge)
Personalization:
to a limited extent (only the order of the content shown might change)

▶ It is not something that has been brought up in complaints to the press council. We cannot say if we have dealt with complaints where news automation and personalization have been used without our knowledge.

---

2 In detail, see my column How self-regulation has responded to news automation (June 3, 2020). **https://www.presscouncils.eu/How-self-regulation-has-responded-to-news-automation**

3 These media councils represent Albania, Austria, Belgium (Flanders, Wallonia), Cyprus, Denmark, Estonia, Finland, Germany, Hungary, Ireland, Luxembourg, Macedonia, Montenegro, Netherlands, Norway, Sweden, Switzerland, Turkey, UK (Impress) and Ukraine.

4 **https://www.jsn.fi/en/lausumat/statement-on-marking-news-automation-and-personalization**

5 **https://www.jsn.fi/en/guidelines_for_journalists**

"most likely not yet". Current applications of news automation are **more modest than one might think on the basis of the hype** around them. However, there is a famous futurological phrase suggesting that *the effect of technological progress is often overestimated in the short run and underestimated in the long run*. Accordingly, as the second takeaway, I will go on to address three aspects that should not be overlooked when news automation is discussed: *data, agency and transparency*.

In addition to the two takeaways above, I will, thirdly, turn my attention to the media councils themselves and critically examine whether they have the knowledge and tools to deal with news automation and whatever else the future brings when it comes. I conclude with the thought that **if it is not the media councils that take the lead, it is going to be someone else**. The result might be that media operators will be forced to make decisions on grounds other than journalistic ones.

## BOX 2

▶ It is vital for journalism to make use of new digital tools. However, I also think that if personalization and news automation are used irresponsibly, there is the danger of surrendering journalistic decision-making power to people outside the editorial office. There might also be a danger when media outlets with poorer financial resources use tools they have not developed themselves. It is important to highlight that the use of news automation and personalization involves a lot of journalistic decisions and one cannot just delegate that decision-making power blindly to a complex robot.

I would also like to emphasize that one of the things that should be expected from journalism is that it gives order to a complex world. What I mean is this: We are constantly provided with a huge amount of information, even if we do not really want it. Organizing that information and making decisions about what is worth emphasizing and what is not should be at the heart of journalism. If the newsfeeds show us what algorithms assume we want to see, journalism is not really carrying out that ordering and organizing task.

▶ The fact that, in the personalization technology, Google and Facebook are operating as black boxes, harms everybody else even though we try to operate in a transparent way.

▶ The use of artificial intelligence must be clear and transparent to the public. The public must have clarity about the choices that lie behind the algorithms. However, at the same time, some form of protection may have to be possible, given the very competitive media landscape. If media design algorithms, they must also be able to protect them. So there are two potentially conflicting needs – transparency and protection – between which the right balance must be found.

▶ The user must be able to choose whether they want personalization or not. The user's option to choose personalization must be presented clearly and simply. Also, someone who does not want personalization should not suffer any disadvantage.

# First takeaway: Full-blooded news automation is still little used

L et's think of news production as **a four-stage cyclical process**, as illustrated in the figure below. First, one searches for and selects a topic and an angle for an article (phase A). After this, one collects source information related to the topic (B) and the text and illustration for the article are produced (C). Finally, the article is published and distributed (D), and the analytics of its success might be used to produce new articles, as represented by the arrow outlined from phase D to phase A.

**Algorithmic tools can facilitate each of these sub-processes.** They can give an alert when something possibly newsworthy has occurred (phase A); facilitate the gathering and then the analysing of source material (B); transcribe interview recordings, proofread manuscripts, edit photos and sync audio and video files together (C); do A/B testing on headings, distribute customized teasers of a published story to various channels, moderate audience comments, and pile up lists of recommended articles that are based on the individual user's past online behaviour (D).
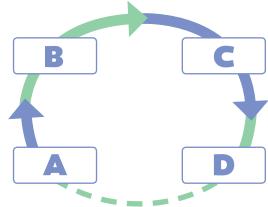


**Figure:** Four-stage cyclical process of news production

It can be argued that in all these examples **there is always a human component involved in the process that ensures that ethical requirements are met** – although the involvement is rather remote with tools like Google Search that are produced outside the newsroom. Nevertheless, someone is more or less in control of the outcome. In full-blooded news automation, on the other hand, **software independently handles the entire cycle** from phase A to phase D. Such solutions can be produced in-house or bought from external software providers.

As already discussed, the perception of media councils is that the software-driven generation of news text that I have labelled as full-blooded news automation is not yet common but, in some countries, that personalization is already more so. To broaden the view, I conducted an online survey with key media operators, "pioneers in the field", whose names were given to me at my request by the media councils. These answers confirm that **we are still dealing with something of relatively rare and limited use**.

The sample (n = 16) is by no means statistically representative, but the answers reflect the general perception that the use of automation in news work is mainly related to **specific visual or numerical parts of an article** (e.g. infographics about stock prices and election results; up-to-date counters for coronavirus infections) **as well as recommendations** based on users' previous online behaviour (e.g. clicks) or input they have entered.

The sample includes media houses (e.g. Jysk Fynske Medier, Mediahuis), news agencies (e.g. Norsk Telegrambyrå AS, Belga) and separate news operators (*Helsingin Sanomat; RTL Nieuws*) in 10 countries.[6] I also approached dozens of the largest media players in the UK, France and Germany, but it proved extremely difficult to get answers from them. According to well-informed guesses on the part of some experts I talked to, this probably reflects the sensitivity of the subject. "In countries such as the UK and Germany, huge number of texts are produced automatically. Since publishers keep quiet for fear of upsetting readers, we do not know to what extent." **Furthermore, news automation may be a field where innovation determines future winners and losers, so nobody wants to reveal their plans.**

In addition to the actual applications already in use, there seem to be clear visions and ongoing development in the field, and many respondents acknowledged **the potential of algorithms (BOX 3)**. All in all, however, **news automation is still a relatively marginal phenomenon, and that is my first reason for arguing that it is still too early to revise the codes of ethics in response to automation.**

**My second argument stems from the fact that existing automation systems are often like showcases** for the software service providers and media operators who have adopted them. It follows that the operators want to ensure that their fancy "newsbots" work properly. It is impossible to imagine and prevent all the mistakes that might arise, but probably even strict self-regulatory guidance would be unable to do any better. In the worst scenario, **excessive self-regulation could hold back development.**

---

6 Belgium, Cyprus, Denmark, Finland, Ireland, Macedonia, the Netherlands, Norway, Sweden and the UK.

## BOX 3

▶ We [a news agency] will continue to produce a vast number of stories based on structured data from financial areas and all the possible statistics we can find. We will also use algorithms to do research, monitor, and probably also produce "half baked" stories that our customers can build on for themselves.

▶ The use of algorithms is not very well established here. The knowledge and the insight into what it can, will or cannot do for us is there. But it is fragmented, and resources are an important reason why we have not yet implemented it in many areas. We recognize the importance of the use of algorithms as well as the fact that their use will continue to grow in the future.

▶ In my opinion, looking at the rate at which a typical newsroom is shrinking in my country, automation of at least some of the news production must increasingly be a tempting option to local publishers.

▶ We want to move to hybrid systems where editorial selection and personalization go hand-in-hand. The editorial team chooses the most important stories and the algorithm is responsible for the order in which they appear, based on users' preferences.

▶ We'll produce a lot more hyper-local content [with the help of automation] – and with personalization we'll be able to target relevant audiences – thereby making our media more relevant to the user.

▶ They will be implemented in all areas where human labour can be feasibly replaced by robots. However, so far there has been more talk than practical solutions.

▶ I believe that artificial intelligence and algorithms will play a much more central role in the future, especially in the areas of recommendation and accessibility, but gradually also in information gathering and the production of content.

# No reason to see monsters where they do not exist

**M**y third argument in support of the view that it is not yet time for major reforms of ethical guidelines relates to **the content that is automated**, and it is chiefly based on my discussions with **four software developers**: Frank Feulner (Chief Business Development Officer / AX Semantics), Sören Karlsson (CEO / United Robots), Jarno Koponen (Head of AI and Personalization, Yle News Lab / Finnish Broadcasting Company YLE) and David Llorente (Founder and CEO / Narrativa).

Today, the state of the art in full-blooded news automation is that algorithms are able to grind news stories on weather, sports, elections, financial reports, breaking traffic incidents, crimes, real-estate prices, natural disasters and so on. The variety of domains is already fascinating. For the domain selection, there are two key criteria.

**The first criterion is the availability of a systematically structured data set** that is numerical or categorical. This means that it consists of numbers or other known values (e.g. names of people), for each of which it is known exactly what each value refers to. The software developers pointed out that their **news automation systems use the same data sets as human journalists use** when covering the above-mentioned topics; data sets that are judged valid by journalists. Looked at the other way round, software does not scrape together rather dubious bits and pieces of information from here and there.

Furthermore, these data sets are often **produced officially, for statistical purposes, not for media**, let alone for news automation. In this respect, it makes little sense to think that authorities would distort their documentation on, for example, police work or air pollution, just to get the software of a certain publisher to write misleading news stories on them. Such misconduct would undermine so much in terms of overall credibility that it will not happen in modern societies, which is precisely where there have been experiments in news automation.

**As to the second criterion: the news topic or event for which the news automation is harnessed must be one that tends to be covered in a certain, predictable way.** This allows the news story to follow a standardized procedure in coverage and text structure; if something completely unexpected happens, the software cannot write about it. In this light, rather than *news stories* or *news articles*, **the outcome of today's news automation systems should rather be called news reports or even *news alerts*.**

As regards personalization, which is perhaps the most controversial aspect of algorithm use, **the current applications in journalism are less dubious or shady than, for example, those in online marketing or e-commerce.** It is unlikely to cause a great deal of harm if you are given a report of the scores of a match from the perspective of either the home or the away team, depending on the geographical location of your IT device.

> It makes little sense to think that authorities would distort their documentation to get the software of a certain publisher to write misleading news stories.

It is also worth reminding that "read more" recommendations, by far the most common form of personalization used in the media industry, **recommend articles on the same topic from the archive of the *same* media** (or media house). They do not bring the reader whatever algorithms happen to find in the bottomless pits of the internet (unlike general search engines), so **the recommended articles meet the quality criteria of that particular media outlet**. Such a service produced by automation is thus comparable to what readers do anyway: they skip the boring articles and read only the ones that interest them.

There may be risks of so-called filter bubbles behind the procedures of personalization, but they must not be exaggerated. So, keeping in mind that we are speaking of journalistic operators, not social media companies that are not committed to any codes of ethics similar to those that govern journalism, there is no reason to see monsters lurking where they do not exist.

# Progress must not be underestimated in the long run

As already mentioned, the media councils have not yet faced ethical problems related to the news automation. Similarly – and not surprisingly – the "pioneering operators" who responded to the survey have not come across any ethical issues in the applications they have already implemented. Nor do their audiences see any, it seems, given that people contacted the editorial office about news automation and personalization "once a month" or less often, mainly asking what user information is collected and for what purpose. However, these pioneers were able to imagine **potential problems (BOX 4** on next page.**).**

Furthermore, the software developers who design solutions for today and tomorrow answered unanimously when I asked about the ethical conflicts they have run into: none. No matter whether, like a news agency, they provide ready-made news texts for newsrooms, or tools with which a newsroom can then automate its news production itself, there have been no practical or ethical issues with the newsrooms, legislators or audience.[7]

**The codes of ethics should only be revised after a clear and lasting change has taken place in the industry** and, as described above, news automation is nowhere near this, yet. The kind of ethical basis that would need to be changed with each inno-

---

7 However, there are also signs of conflicts in the research literature, when software developers and journalists try to reconcile the worlds of technology and journalism, see, e.g., Dierickx, L. (2019) Information automatisée et nouveaux acteurs des processus journalistiques. *Sur le journalisme, About journalism, Sobre jornalismo* 8(2): 154–167. Available at **http://www.surle-journalisme.kinghost.net/rev/index.php/slj/article/view/408**

vation and new kind of experimentation would actually seem rather unstable. However, this is not where the story ends – this is where it begins. Let's revisit the futurological phrase, and especially its second half: **while the effect of technological progress is often overestimated in the short run,** *it is underestimated in the long run.*

**Things are happening in the field of journalism, and innovations also spill over from nearby industries,** not least because for many software development companies, journalism is just one object of automation; personalization, for example, is already very common in online marketing. Possible future developments were also highlighted in the research literature used as a basis for this report (listed at the end of the report) as well as in discussions with other researchers.[8]

In addition, what comes next also draws on two research projects in which I have been involved. In the *Immersive Automation project* (2017–2018), researchers and media companies worked together with data scientists, linguists and journalists to create a roadmap and a demonstration of a future news ecosystem based on algorithmic journalism. In the *Targeted Journalism* project (2020–2022), researchers are examining the ways, goals and the future of personalization.

---

8 Laurence Dierickx (Scientific collaborator at the ReSIC research center, Université Libre de Bruxelles), Carl-Gustav Lindén (Associate Professor [Data Journalism], University of Bergen) and Henrik Rydenfelt (Adjunct Professor, University of Helsinki).

# Second takeaway:
# Be prepared for news automation

News automation can mean many things, but in this report I mainly refer to **Natural Language Generation (NLG). It is a process of translating digitally coded data into human language, from numbers to text.** This process needs, to put it simply, two things: a structured and verified data set about the topic in question, and guiding principles on how to write about it.

In general, NLG solutions vary from systems based on rules and templates written by humans to those based on machine learning methods. While the latter approach, data-driven machine learning, is in many ways attractive – and already in use in a side role in news automation – it is difficult to have full control over its actions. **Journalism requires high accuracy and accountability, and that makes the use of machine learning systems problematic as it increases uncertainty and risks of failure.**

One such problem is "overfitting", where an NLG system fails to generalize to cases outside the training data. For example, a system for producing restaurant reviews might learn from training data that restaurants of a certain type are almost always inexpensive. This results in the system generating reviews where all restaurants of that type are characterized as inexpensive, regardless of their real pricing.

Despite the challenges, the role of machine learning is increasing and making news automation more useful. For example, one software developer I interviewed said they have automated the market stories of a major media operator. For this, they received a teaching corpus of 10,000 human-written market stories to learn the style, structure and importance of certain features. In less than two weeks, they had a library of templates. Machines may make mistakes, he told me, and that was why people fine-tuned the sentences. "Often, however, the errors are due to the teaching corpus. Perhaps some of the stories in the corpus are not valid on that topic or contain ambiguous expressions."

When one lets one's imagination run freely, beyond the bounds of reality, one often ends up indulging and embellishing the idea that by comparing different data sets with high computational power, algorithms reveal hidden correlations and create analytical insights. The software developers, however, were not talking about such scenarios. "You won't find an editorial that would have been done with our automation", one of them told me, "but what you will find is data-driven and highly repetitive content that scales to many dimensions".

On this basis, **the realistic vision, in the short run, is that news automation systems will grind report-like news stories also in the future.** This, in turn, resonates nicely with the oft-heard idea that news automation frees up journalists from the monotonous, routine tasks and gives them more time for the more challenging tasks.

**Despite the challenges, the role of machine learning is increasing and making news automation more useful.**

However, this too seems to be an exaggeration: the fact seems rather to be that **news automation generates news stories in mass that would not otherwise be done at all.**

After all, the most important limitation that holds up the progress of news automation seems to be that **algorithms rely, so to speak, only on a quantified version of reality, thus making *data* the principal bottleneck.** This is something considered both in the research literature and the interviews, and it is therefore the first of the three aspects I have identified that should not be overlooked when the future of news automation is discussed.

# Data: Quality of output depends on input

*Garbage in, garbage out*, as the well-known saying goes. In other words, **the quality of output of an automated journalistic system is highly dependent on the data that is fed into it.** This quality can be described and evaluated through five V's: volume, velocity, veracity, variety and value.[9]

Let's start with *volume* and *velocity*. An example of high-volume data is elections, which provide a lot of data but on a one-off basis, so the velocity is low. In contrast, a city's temperature readings exemplify low volume, as they provide little data at a time, but their velocity is high because these readings can be done very often, thus producing a steady stream of data. An example of high volume and high velocity would then be detailed meteorological information produced by all of a country's weather stations.

However, data produced by a broken weather sensor is worthless. Therefore *veracity* is naturally key in all journalism. In order to use a data set, we need to know who produced it and for what purpose, and how its accuracy is ensured now and in the future. We also need to see if there could be intentional or unintentional distortion in the data set, such as human bias, or even the possibility of manipulation by a third party.

The last two V's are the most crucial ones: **real added *value* – and with it the breakthrough in news automation – will only be created through *variation*.** News stories of temperatures or elections results are hardly something that will make the audience change their habits or whip out their credit card and thus alleviate the financial distress of the media. So it is not enough to automate the topics from which media operators happen to obtain suitable data sets. There needs to be a change of direction.

**Suitable data sets need to be produced on the domains that media operators want and need to address.** For example, if news automation could convert written parliamentary minutes, audio recordings of debates – perhaps including the visual dimension, too – and the social media posts of the candidates in an election into

applicable data sets, the audience's decision as to who to vote for could be better informed – and the automated media story would truly make a difference.

Exploring the possibilities of utilizing new types of data sets is an ongoing topic of research. Such data sets cannot usually be produced within the newsroom, nor would it be appropriate. It requires **the collaboration of experts from many different fields,** which leads us to discuss the *agency* of producing algorithms and datasets.

# Agency: Who makes the algorithm?

In the ongoing discussions on the ethics of automated journalism, *agency* and thus responsibility are only rarely attributed to algorithms and news robots. In other words, the prevailing view is that newsroom processes are controlled by human beings and responsibility is assigned to them.[10] However, with the rise of news automation the question is, who are the *human beings*?

Writing algorithms and producing data sets requires a kind of technical expertise that journalists and other traditional newsroom workers rarely have. This means that **media operators need new types of professionals** who can approach journalistic processes from the perspective of computational thinking by breaking down a complex problem, focusing on the crucial parts, developing a step-by-step solution and, finally, coding this solution into an algorithm.

The challenge is that the professionals with computational thinking, let's call them software developers, are probably not trained in journalism but in computational linguistics, engineering or in data science, and they are socialized into the world of technology. **Despite the fact that the worlds of both technology and journalism serve society, their responsibilities are directed differently:** software developers are accountable first of all to their employers, and they are not responsible for any potential misuse of their technology, while journalist are accountable first of all to their audiences, and they are responsible for what they produce.

Therefore the software developers – whether an integral part of the editorial team or external software designers – need to internalize their role as part of the editorial production chain and adhere to a journalistic value base when making independent decisions that influence journalistic content. Here, **communication becomes a crucial part of software development.** This requires a shared language and a dialogue on equal terms between journalists and newcomers that will, I argue, benefit them both in various ways.

Firstly, when we talk about automation in a journalistic context, as in this report, the talk forefronts the new kind of expertise that is needed. At the same time, the expertise involved in being a journalist is taken for granted as naturally as the air that journalists breathe, and is easily ignored. Journalistic expertise is often based on practice and is mostly intangible. When this expertise is structured and explained to

---

10 As Henrik Rydenfelt shows in his forthcoming paper (for detailed reference, see *Suggested readings*), confidence in (human) editorial control and its persistence is strong – even unnecessarily strong?

software developers, it will also become more visible. This is likely to **raise the profile of journalism** as a profession.

Secondly, in order for software developers to adhere to the journalistic set of values, this value base needs to be discussed explicitly and analytically. In practice, this will force journalists as well as media councils **to explain and define those vague concepts such as "journalistic decision-making power",** "journalistic principles" and "news value" that they often use so gladly. This points to *transparency*, the third aspect that is important in a discussion of news automation.

# Transparency: To increase the audience's trust

In the past, journalists, by virtue of their profession, had quite exclusive rights both to sources of information and to distribution channels, and that placed them in a powerful position to support and steer public discussion. With the internet, however, many people now have similar access to information, and this has opened people's eyes to the realization that **there is no *objectivity***; instead, making news is full of deliberate choices. And as the audience's illusions about objectivity and impartiality have crumbled, so too has public trust in journalistic media.

*Transparency* about journalistic processes has been suggested as a solution to break this unfortunate cycle and **restore a trusting relationship between the news media and their audience**. When we talk about the use of automation and personalization, which might sound a bit scary to some, the need for transparency is obvious. At the same time, **transparency would be a way for journalistic media to differentiate themselves from the social media giants**, whose use of algorithms has raised many doubts.

> As the audience's illusions about objectivity and impartiality have crumbled, so too has public trust in journalistic media.

According to the research literature, there are three important aspects to transparency.[11]

Firstly, transparency can be enhanced by highlighting that the content is produced algorithmically. In addition, it makes sense to point out in more detail **for what purpose the automation system in question is designed, and by whom**. In a brand-new book, *New Laws* on robotics (Harvard University Press, 2020), Frank Pasquale presents "four new laws of robotics", the fourth of which (p. 11) is that *Robotic Systems and AI must always indicate the identity of their creator(s), controller(s) and owner(s)*.

Care should be taken, however, to ensure that the pursuit of transparency does not become just a list of technical information, let alone routine pop-up windows. That would be another example of how a good intention can turn against its purpose – just

---

[11] This section draws on a forthcoming paper, *Dataa näkyvissä: Läpinäkyvyys algoritmien ja datan journalistisessa hyödyntämisessä* [Transparency in the journalistic utilization of algorithms and data], I have co-authored with Tuukka Lehtiniemi and Henrik Rydenfelt.

think about the cookie consents that pop up on every website you visit (when using the internet in Europe).

Secondly, there is a need to throw light on processes. How are the data sets selected or produced, why are they suitable for the purpose at hand, and how does the algorithm utilize them in automation? Note: just opening the bonnet is not enough in itself to create understanding of how the car engine works, and similarly, **news automation needs to be explained in such a way that the audience really begins to grasp the "automagic" taking place behind the scenes.**

Thirdly, transparency is enhanced by public involvement. This may relate, for example, to data journalism articles, such as one[12] in which the reader enters information about how much money they spend and receives a calculation of how tax increases will affect their expenses. This is an active and fairly public interaction that in itself makes journalism transparent. However, there are also so-called interactions of which the audience is not even aware, and that is especially true with personalization. In the aftermath of the Cambridge Analytica affair[13] and other scandals, which have raised people's awareness of the tracking issue and possible invasions of personal privacy, **readers' and viewers' trust should be bolstered by telling them how and what data are collected about their online behaviour, and how and for what purpose these data are used and for what goals.**

Worth pointing out that it also makes sense to tell readers that a news site is ***not personalized***. Otherwise if, for example, someone is planning to buy a new car and has mentioned it here and there, and then they read an article about shopping that happens to be illustrated with photos of their favourite car, they might wonder if they are being spied on.

Finally, the role of journalism is to describe and explain the world, and today what is perceived as reality is shaped significantly by data and algorithms. **Thus, in opening up its own use of data and algorithms, journalism will also be opening up the data-driven world in which we live.**

---

12 **https://www.hs.fi/politiikka/art-2000006242236.html** (in Finnish)
13 **https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal**
    (read Nov. 29, 2020)

# Some experimental support for Takeaways 1 and 2

**A YEAR AGO,** Finland's Council for Mass Media published a pioneering statement on the journalistic use of algorithms (*Statement on marking news automation and personalization*). It did so despite the fact that, according to a background survey on the Finnish mediascape, **the use of news automation was experimental at that time, and personalization only a little more common**. Additionally, the council had received only one complaint related to algorithms. Nevertheless, the council decided to act proactively. **These empirical observations by Finland's Council for Mass Media contribute to my first takeaway, that creating self-regulation for the journalistic use of algorithms is not yet an especially topical issue.**

Furthermore, the Finnish *Guidelines for Journalists* call for thorough source criticism on the part of journalists, and this can be seen to cover issues related to *data* as well. The statement supplementing the Guidelines then covers the issues of *transparency* and *agency*: with regard to transparency, the Statement makes recommendations on how to mark news automation and personalization. With regard to agency, the Statement defines the use of algorithmic tools as part of journalistic work and calls for digital service developers also to adhere to the Guidelines.

The Statement received a lot of publicity. The *Finnish Broadcasting Company Yle* published a thorough – but still easy-to-read – report on how it uses news automation and personalization, and *Helsingin Sanomat*, a leading newspaper, said that it would carefully review its procedures to make sure that they meet the requirements.[1] However, **what supports my second takeaway – that data, agency and transparency are probably relevant, but will only really be so in the future – is that, firstly, the Statement has not provoked significant action or debate in the Finnish media industry during its first year, meaning that no one felt an urgent need for that, and, secondly, no complaints have been lodged with the council on the grounds mentioned in the Statement.** However, it should also be noted that there is no general debate on the wider applications of algorithmic decision-making systems in society, despite the ever-increasing diversity of their use.

1 See, **https://yle.fi/aihe/sivu/yleisradio/ylen-palvelujen-personointi-ja-uutis-automaatio** and **https://suomenlehdisto.fi/jsn-ohjeistaa-kertomaan-algorit-mien-kaytosta-meidan-pitaa-olla-avoimempia-kuin-muut/** (both in Finnish).

# Third takeaway: The councils' procedures must be reconsidered

The third takeaway of this report draws attention to the work of the councils. As I have already mentioned, news automation that covers the entire editorial process from topic selection and information collection to text production and publishing is still rare. However, one of its features, personalization, which is sometimes considered controversial because it might create filter bubbles, is already pretty common. **Personalization is also something that media operators described as more attractive to develop in the short term than full-blooded, text-oriented news automation.**

It would be tempting to say that news automation in general and personalization in particular work ethically and without problems, as almost no complaints have been lodged about these issues with European media councils. However, this might be too simple an answer. Instead, could it be that **the lack of complaints is rather a symptom of problems with the functioning of the media councils?** This doubt has grown in my mind with my membership of the Finnish Media Council (2017–2019), and according to the excellent study conducted by Raymond A. Harder (For detailed reference, see *Suggested readings*) as part of the project *Media Councils in the Digital Age*, it might apply elsewhere in Europe, too.

To begin with, **the rules of media councils are formulated positively** in the sense that they only allow people and organizations to complain about things that have already been published or done. One also needs to identify a well-defined subject for the complaint, often a single piece of incorrect information, and **one cannot complain about something that is missing in a report**. Furthermore, what might be considered particularly **alarming** is the fact that many councils (11/28 investigated in the above-mentioned study) do not allow complaints about articles that do not personally involve the person complaining. In other words, these **councils require that you have a personal stake** in the article that you are complaining about– the article needs to be about you, your organization, your ethnic group, etc.

News automation and personalization hardly fit into such specs. They do not target an individual actor as required by the guidelines, neither the process nor the product are well-defined, and they take place behind the scenes, making it difficult for the audience to know about them. **The bottom line is that even if news automation and personalization did have ethical or other issues, it would be next to impossible for any member of the audience to lodge a complaint that would end up in the council.**

My third takeaway relates to the procedures of complaining. In order to meet the self-regulatory needs of news automation – and at the same time also the needs of

> **What might be considered alarming is the fact that many councils do not allow complaints about articles that do not personally involve the person complaining.**

other innovations related to digitalization, such as web-based audience interaction and the journalistic use of social media – **the media councils should ensure that they offer a realistic opportunity for the audience to complain.** However, responsibility should not rest so firmly on the audience alone. Instead, councils should be sensitive to silent signals and act on their own initiative. **Almost all councils have the possibility of proactively raising for their own consideration and for public debate matters of media ethics, but most of them rarely, if ever, use this possibility.** In addition to the problem of resources, the councils themselves referred to the fear of becoming perceived as the "media police".

# Conclusion

In this report, I have discussed how the media councils should respond to the fact that news automation – i.e. NLG systems writing journalistic articles from start to finish without much human input – is already here, and it will be much more powerfully so in the future. Based on what constitutes the present state of the art in news automation and what ethical considerations it gives rise to, this report has presented three key takeaways.

Firstly, I have argued that there is as yet no urgent need for self-regulatory guidance on news automation. News automation systems are still rare and are more like showcases than profitable everyday tools. Secondly, I identified three aspects – namely *data, agency,* and *transparency* – that should not be overlooked in the discussion of news automation. They will become very relevant sooner rather than later.

Thirdly, I argued that media councils should critically revisit their complaints procedures and, most importantly, act proactively in relation to the changing media landscape. **The fact is that if the media councils do not take the lead on this, someone else will do so, and whoever that is, whether it is national legislators, the EU, or platform companies, this could put the freedom of the press in jeopardy.**

# Suggested Readings

► Algorithms, automation, and news (2019) *Digital Journalism* (special issue) 7(8). **https://doi.org/10.1080/21670811.2019.1685395**

► Beckett, Charlie (2019) New powers, new responsibilities. A global survey of journalism and artificial intelligence. *The Journalism AI*. Available at **https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities**

► Caswell, David & Dörr, Konstantin (2017) Automated Journalism 2.0: Event-driven narratives. *Journalism Practice* 12(4): 477–496. **https://doi.org/10.1080/17512786.2017.1320773**

► Diakopoulos, Nicholas (2019) *Automating the news. How algorithms are rewriting the media*. Cambridge: Harvard University Press.

► Dörr, Konstantin & Hollnbuchner, Katharina (2016) Ethical challenges of algorithmic journalism. *Digital Journalism* 5(4): 404–419. **https://doi.org/10.1080/21670811.2016.1167612**

► Haapanen, Lauri (2020) Problematising the restoration of trust through transparency: Focusing on quoting. *Journalism* (published online before print). **https://doi.org/10.1177/1464884920934236**

► Haapanen, Lauri & Leppänen, Leo (2020) Recycling a genre for news automation: The production of Valtteri the Election Bot. *AILA Review* 33: 67–85. **https://doi.org/10.1075/aila.00030.haa**

► Harder, Raymond A. (2020) Media councils in the digital age: An inquiry into the practices of media self-regulatory bodies in the media landscape of today. Vereniging van de Raad voor de Journalistiek. Available at **https://www.presscouncils.eu/userfiles/files/Media%20Councils%20in%20the%20Digital%20Age.pdf**

► Lindén, Carl-Gustav (2017) Decades of automation in the newsroom. *Digital Journalism* 5(2): 123–140. **https://doi.org/10.1080/21670811.2016.1160791**

► Lindén, Carl-Gustav & Tuulonen, Hanna (eds) with Asta Bäck, Nicholas Diakopoulos, Lauri Haapanen, Leo Leppänen, Magnus Melin, Myriam Munezero, Stefanie Sirén-Heikel, Caj Södergård, Hannu Toivonen (2019) *News Automation – The rewards, risks and realities of 'machine journalism'*. WAN-IFRA guide to the field. The World Association of Newspapers and News Publishers WAN-IFRA. Available at **http://immersiveautomation.com/wp-content/uploads/2019/06/WAN-IFRA_News_Automation-FINAL.pdf**

► Rydenfelt, Henrik (forthcoming) Transforming media agency? Approaches to automation in Finnish legacy media. [Journal article manuscript.]