Algorithmic Journalism and Its Impacts on Work

Ayse Dalgali
Syracuse University
School of Information Studies
ayocal@syr.edu

Kevin Crowston
Syracuse University
School of Information Studies
crowston@syr.edu

Abstract

In the artificial intelligence era, algorithmic journalists can produce news reports in natural language from structured data thanks to natural language generation (NLG) algorithms. This paper presents several algorithmic content generation models and discusses the impacts of algorithmic journalism on work based on O*NET data and within a framework consisting of three levels: replacing tasks of journalists, increasing efficiency, and developing new capabilities within journalism. We also discuss journalism ethics associated with the growth of the algorithmic journalism field. The findings indicate that algorithmic journalism technology may lead to some changes in certain tasks of journalists and some shifts in journalism practice by enabling individuals to produce their own stories. This paper may contribute to an initial understanding of how algorithmic news is created, how algorithmic journalism technology impacts journalists’ work, and what type of ethical issues brings with it.

1. Introduction

In the present era, artificial intelligence (AI) applications address diverse tasks including image recognition, machine translation, and guidance for automated vehicles. And in journalism, they are now used for much harder cognitive tasks, such as content generation, content editing, combining databases with editor-created story templates to generate stories, and editing reports, all of which journalists used to perform before. Companies such as Arria, AX Semantics, Retresco, Automated Insights, Narrative Science, Associated Press, and Gannett [1,2] utilize algorithmic journalism. For instance, the Associated Press generates around 3,700 earnings reports on US and Canadian companies using AI technologies [25]. Narrative Science produces algorithmic news from economic indicators and game reports [22].

Yahoo uses Wordsmith to prepare texts for fantasy sports games [16].

To date, emerging technologies have shaped and affected journalism and news reports’ production traditions [11]. Algorithms get involved in the processes of generating reports may have further impacts on current journalistic work routines or on journalists’ specific tasks or roles. These possible shifts can be observed in journalism practice may bring with it new measures to be taken to protect ethical values and responsibilities attributed to journalism [11]. In this paper, we define algorithmic news, discuss the underlying technology, and the impacts on the work of journalists. More specifically, we examine the journalism tasks obtained from O*NET to investigate which tasks can be overtaken by the algorithms and in which roles or tasks of journalism shifts may be occurring. Additionally, we discuss journalism ethics associated with the growth of the algorithmic journalism. Thus, the paper examines two research questions:

RQ1: Which tasks of journalists can be overtaken by algorithmic journalism? In which roles or tasks of journalism shifts may be observed?

RQ2: What ethical challenges may emerge as a result of algorithmic journalism?

To describe the use of AI technologies in journalism, different terms are used such as AI journalism, automated journalism, and robot journalism [6]. In [17], the term automated journalism is defined as “any process or system of news production under the control of media or electronic devices, with little or no external influence” (p. 6). Ref[4] also uses the term automated journalism and the definition used in that source is cited by many articles: “journalism in which a program turns data into a news narrative, made possible with limited—or even zero—human input” (p. 416). On the other hand, the term robot journalism might be understood to mean “physical robots” that are envisioned as replacing news-casters in newsrooms [20], making it less appropriate for
our purposes. In this paper, we use the term algorithmic journalism, following Dörr's definition [10]:

the (semi)-automated process of NLG (natural language generation) by the selection of electronic data from private or public databases (input), the assignment of relevance of pre-selected or non-selected data characteristics, the processing and structuring of the relevant data-sets to a semantic structure (throughput), and the publishing of the final text on an online or offline platform with a certain reach (output). (p.702).

To define the news created by these processes, we will use the term algorithmic news, and to define the algorithms that generate algorithmic news, we will use the term algorithmic journalists. Algorithmic news refers to news reports generated by algorithmic journalism. According to Dörr's definition, the output is described as the final text published, and the throughput is described as a semantic structure. Algorithmic news reports are publishable texts that have a semantic structure created by algorithms from data.

The remaining of the paper is organized as follows. We first review the studies describing the technologies behind the algorithmic journalism; evaluating the changes in the roles assigned to journalists because of algorithmic journalism; and discussing the ethical challenges that algorithm journalism brings with it. We then analyze the journalists’ tasks presented in O*NET to explore which tasks can be overtaken by algorithms or which changes can be observed in which roles of journalists. Next, we discuss our findings and then conclude the paper.

2. Related Work

2.1. The Technologies behind the Algorithmic Journalism

The main technology used in algorithmic journalism [8]–[10] is natural language generation (NLG), a subfield of natural language processing (NLP), which describes a software process in which structured data are converted into human (natural) language. Other technologies may also be used to generate news content. For example, Narrative Science and Automated Insights graphics and pictures to their generated texts [10]. These additions are different from NLG [10], but these visualization tools, which improve the diversity of the content, may help to generate more enjoyable and attractive news. In many applications, human intervention is still needed; that is, the process may not be fully automated.

Dörr’s algorithmic journalism model, an input-throughput-output (I-T-O) model (shown in Figure 1), is based on the model proposed by Latzer, Just, and Saurwein. Dörr notes that algorithms behind algorithmic journalism apply the rules of NLG [10]. Hence, in the I-T-O algorithmic journalism model, NLG also carries out algorithmic selection. The process of generating news algorithmically starts with a database, such as sports, financial, weather, or traffic data (input). Next, the data are converted according to predefined linguistic and statistical rules (throughput) into a text (output) in natural language [10].

In Dörr’s model, content is generated in three stages: (1) text planning, (2) document structuring, and (3) structure realization. The purpose of content determination (input level) is to decide which information is helpful to the user or important for the expected output. The input to the text planner (i.e., structured data) is the input to the entire process of NLG. This structured data is accessible through public application programming interfaces (APIs) or through private databases (e.g., commercial data). Because algorithmic news, based on NLG, consists of content related texts, to generate these texts, specific codes, rules, and dictionaries are used and adapted. Hence, NLG operates based on pre-set special rules regarding the linguistic creation process and criteria for identifying and selecting facts in the data to be processed and transformed into natural language [10].

![Figure 1. I-T-O model of algorithmic journalism](image-url)

In the planning stage, at the input level (request), features such as text length, journalistic genre, tonality, and the time and place of publication are determined. In the throughputs stage, based on the criteria arrived at in the planning stage, the NLG algorithm selects components from the data set, and aggregates and assigns relevance to them. After this process, the algorithm identifies the linguistic structures (words, syntax, sentences) to be used to achieve the desired information, the forms of words to use, and their order of appearance. More specifically, the NLG algorithm makes lexical choices and decides which content and words should be used to explain domain concepts; makes syntactic
choices and decides which syntactic structures should be used in generating sentences; and aggregates data and decides how many messages should be included in each sentence. After the text in natural language (i.e., the output result) is generated, humans intervene in the process through the feedback loops to optimize this generation until the intended result is accomplished. Finally, after the text-generation process is completed, the texts are often published automatically on online or offline news outlets [10].

Although Dörr’s definition includes the term semantic structure, previous NLG models that generate news were only simple descriptions of routine sports and financial news [10]. The I-T-O model and other previous NLG algorithms for generating news did not contain semantic features. Therefore, they can produce only short, simple descriptive news in limited domains, rather than more complex news like that generated by humans, such as event-driven narratives [5]. Caswell and Dörr [5] relate this problem to the absence of semantic elements in data and the absence of appropriate data models (methods for processing data) from the production of more complex algorithmic news. The common previous method for algorithmic journalism used trees and templates, for example, if the data field 1 is X, then write Y, which creates story templates. After these story templates are created, when new data are collected, the templates are filled out by the new data based on the trees.

To address the limitations of these approaches, Caswell and Dörr [5] proposed a model that uses semantic features to generate event-driven narratives. Caswell and Dörr’s model is based on a “story database.” Journalists enter events and narratives into this database, which uses the semantics, or meanings, of journalistic events to categorize news stories in the form of structured data. This model merges NLG with structured data that represent stories semantically in a story database to generate complex journalistic narratives as illustrated in Figure 2.

To contribute to the story database, journalists enter data according to the semantic features of actual news reports. For example, if the news report is related to commerce, it is stored under the “commerce” semantic frame, which entails certain roles, including buyer, seller, etc., and certain actions, including buying, selling, paying, etc. In Caswell and Dörr’s model, first groups of related semantic frames from structured stories are chosen. Then, for each group of semantic frames, an appropriate template created by Wordsmith with blanks is filled out with the relevant structured data obtained from the story database, based on the semantic features and context of the news report to be generated. Then, these completed templates (i.e., text blocks) are combined to generate a complete news report.

In 2015, Wordsmith, an artificial writer (i.e., and algorithmic journalist) using NLP, developed by Automated Insights, one of the leading commercial providers of NLG technologies, became available [23]. Wordsmith branches paths (i.e., creates trees) by adding words, sections, or phrases, or modifying or removing them [23]. A Wordsmith user enters data, such as criminal records, and then Wordsmith builds branches around that data. This process constitutes a story structure, which is used as a template for numerous further articles. A sample report concerning crime trends produced and shared by Automated Insights is shown in Figure 3 [23].

![Figure 2: Generating algorithmic news from structured stories [5].](image)

![Figure 3: Example crime report created by Automated Insights [23].](image)
2.2. Changes in the Roles Assigned to Journalists Emerging with Algorithmic Journalism

Previous work on the impacts of automation has mainly focused on possible job losses emerged with algorithmic journalism [3, 6, 18, 19]. A few studies examine the changes in the roles of journalism that algorithmic journalism brings with it. These include Schapals and Porlezza’s study [24] that investigates how journalists are influenced by the technologies attributed to algorithmic journalism, and to what degree journalistic roles are challenged or advanced as a result of algorithmic journalism technologies. Drawing on semi-structured interviews with journalists from German newsrooms, the authors found that journalists commonly view the advantages of algorithmic journalism with the idea that it will supplement journalists rather than replacing them or competing with them [24]. They believe that algorithms assist them in their daily news work.

Schapals and Porlezza did not analyze the specific roles of journalists. They rather examined general more comprehensive roles that cover the specific roles and constitute journalism profession, e.g., normative roles, narrated performance, public service, etc. [24]. The journalists interviewed in [24] advocate that journalism is an individual craft requiring creativity; hence, narrated performance of journalism cannot be beaten by algorithms, thus normative roles (e.g., generating creative news reports as a product of craft) expected from journalists still be present, even strengthened [24]. The journalists envisioned themselves still as authority in their public service role, such as presenting and spreading information and being ‘watchdogs’ over society.

Bucher’s [3] study also highlights the supportive role of algorithms. The study investigates how computational artifacts are used in newsrooms through interviews with Swedish journalists. As analogous to the mentioned findings of previous studies, Bucher’s [3] study found that involving computational processes in the journalists’ work supplement journalists rather than replacing them. A master thesis [20] investigated the deployment of algorithmic news in Norwegian newsrooms through the interviews with 11 journalists using algorithmic journalism in Norway, Sweden, and Germany. The thesis demonstrated similar findings to the findings of prior research: the journalism profession will still preserve its strength; algorithms are suitable for completing certain tasks, such as mundane and repetitive tasks; therefore, with their use, journalists can devote their time for more creative and challenging work, such as interpreting the algorithmic news and gathering data by the means of interviews.

2.3. Ethical Challenges of Algorithm Journalism

As for journalistic ethics, there are concerns regarding the potential emergence of problems concerning transparency, privacy, bias, etc. Another main concern is that if ethical rules are violated, who will be responsible for this violation? Thurman et al. [25] examined what journalists think about algorithmic journalism by conducting workshops and semi-structured interviews with ten professional journalists from the BBC, CNN, and Thomson Reuters. A journalist stated that if the readers do not know how the news is generated, whether by a human journalist or by an algorithmic journalist, this may cause problems, such as not being able to determine who or what should be credited and held responsible for the output [25]. There could be violations of policy-related rules relating to privacy, security, misinformation, disinformation, fake news, etc., or mistakes contained in news stories due to errors in data. This question concerns liability. To solve this problem, Diakopoulos recommends providing algorithmic transparency by providing data, the model used to process the data, and the results, including errors [9]. However, if this transparency reveals source code and a detailed-enough methodology, may be used to violate anonymity or privacy. Thus, Diakopoulos suggests that transparency should be provided to some extent, not full transparency nor lack of transparency, but balanced transparency for appropriate people [9].

Concerning bias, some journalists think that algorithmic journalism reduces bias that stems from human subjectivity [25]. Nevertheless, some of them argue that it might increase bias because of possible data manipulation by humans desiring to generate biased news by using biased data or biased models that process the data in automatic news generation. Regarding verification, although algorithmic journalism can eliminate human error, it hinders readers from verifying whether the data source is valid and reliable [25].

Since the algorithmic journalism topic is still a new topic, the questions asking about which ethical challenges may emerge as the result of algorithmic journalism, which tasks of journalists can be taken on by algorithmic journalism; in which roles or tasks of journalists shifts may be observed; have not been answered sufficiently yet. The answers to these questions can offer insights with regards to how journalism practice may be shaped by the algorithmic journalism, in particular, what kind of shifts may be observed in journalists’ tasks and roles as a result of algorithmic journalism, and how ethical issues can be addressed properly. Exploring algorithmic journalism’s these impacts is important to enhance its potential advantages.
3. Method

To explore how algorithmic journalism may change current journalism practice, we conduct a conceptual task analysis. We obtained data to be used in the task analysis from the Occupational Information Network (O*NET OnLine). The tasks we analyzed specifically are those assigned to reporters and correspondents (27-3022.00 - Reporters and Correspondents) [27].

The U.S. Department of Labor gathers data for O*NET from the surveys administered to randomly selected U.S. workers in each occupation [7]. The first survey for O*NET was conducted in 1998 and is updated on the website periodically [7]. In this study, we used the latest version, updated in 2020. The O*NET surveys ask various questions about tasks, abilities, technology skills, knowledge, and work activities required for each specific occupation.

O*NET data usage has been recognized by many researchers working on occupation-related studies, e.g., [7, 12, 14, 15, 21]. These studies mainly aim at understanding the properties of work, specifically the properties of a certain job or the requirements for individuals from this job. For example, in [15], job information obtained from O*NET is analyzed to make predictions for requirements with regards to relevant jobs. Additionally, studies introducing O*NET and highlighting this data source’s importance for occupation or job-related research exist, e.g., [13, 21].

In conceptual task analysis, we examined all the 27 tasks attributed to journalists by a conceptual analysis perspective. More specifically, we selected the tasks related to generating news reports based on their relevance to each other, grouped them, and formed new task groups that reflect the combination of the selected original tasks. We excluded the tasks that are not appropriate for fitting to a group. Analyzing these 27 tasks and selecting the ones relevant to generating news, we drew the data flow diagram to demonstrate journalists’ workflow as shown in Figure 4. In the Findings section, based on the tasks and task groups we created, we discuss the possible changes or impacts in the roles and tasks of journalists associated with algorithmic journalism.

4. Findings

4.1. Tasks in Journalism Work

In this section, to answer the first research question, we discuss the impacts of algorithmic journalism on work on the grounds of the conceptual task analysis. O*NET attributes these 27 tasks to journalists, especially to reporters and correspondents [27]:

- T.1. Receive assignments or evaluate leads or tips to develop story ideas.
- T.2. Research a story’s background information to provide complete and accurate information.
- T.3. Arrange interviews with people who can provide information about a story.
- T.4. Establish and maintain relationships with individuals who are credible sources of information.
- T.5. Report news stories for publication or broadcast, describing the background and details of events.
- T.6. Gather information about events through research, interviews, experience, or attendance at political, news, sports, artistic, social, or other functions.
- T.7. Revise work to meet editorial approval or to fit time or space requirements.
- T.8. Review and evaluate notes taken about news events to isolate pertinent facts and details.
- T.9. Investigate breaking news developments, such as disasters, crimes, or human-interest stories.
- T.10. Review written, audio, or video copy and correct errors in content, grammar, or punctuation, following prescribed editorial style and formatting guidelines.
- T.11. Report on specialized fields such as medicine, green technology, environmental issues, science, politics, sports, arts, consumer affairs, business, religion, crime, or education.
- T.12. Determine a published or broadcasted story’s emphasis, length, and format and organize material accordingly.
- T.13. Transmit news stories or reporting information from remote locations, using equipment such as satellite phones, telephones, fax machines, or modems.
- T.14. Check reference materials, such as books, news files, or public records, to obtain relevant facts.
- T.15. Discuss issues with editors to establish priorities or positions.
- T.16. Photograph or videotape news events.
- T.17. Take pictures or video and process them for inclusion in a story.
- T.18. Present live or recorded commentary via broadcast media.
- T.19. Conduct taped or filmed interviews or narratives.
- T.20. Develop ideas or material for columns or commentaries by analyzing and interpreting news, current issues, or personal experiences.
- T.21. Communicate with readers, viewers, advertisers, or the general public via mail, email, or telephone.
- T.22. Write online blog entries that address news developments or offer additional information, opinions, or commentary on news events.
- T.23. Assign stories to other reporters or duties to production staff.
- T.24. Write columns, editorials, commentaries, or reviews that interpret events or offer opinions.
**T.25.** Edit or assist in editing videos for broadcast.

**T.26.** Write reviews of literary, musical, or other artwork, based on knowledge, judgment, or experience.

**T.27.** Participate in community events, make public appearances, or conduct community service.

As briefly mentioned in the Methods section, to draw the data flow diagram, we combined relevant tasks. For example, we combined the tasks T.2, T.4, T.6, and T.14 and created the step “Collect information about stories/events.”

The data flow diagram begins with the T.1, that is, receiving the assignment which is followed by collecting information (the combination of T.2, T.4, T.6, and T.14). It is followed by the task of T.12 – determining a story’s emphasis, length, and format and organizing material accordingly – is one of the first steps for generating news reports as shown in the data flow diagram; and its outputs, “content determination” is input to the further steps which are “arrange interviews with people who can provide information about a story” and “collect pictures or video and process them for inclusion in a story.”

After generating the news reports, journalists perform the tasks of T.7, T.8, and T.10 to review the news reports. We combined these tasks as “review the reports.” As final steps, based on the reviewed reports, journalists perform task T.13, which is “transmit news stories or reporting information from remote locations, using equipment such as satellite phones, telephones, fax machines, or modems.”

The diagram does not include the tasks T.9, T.15, T.18, T.20, T.21, T.23, and T.27, because they are separate tasks that are not directly related to the processes of generating news reports.

---

**4.2. Impacts of Algorithmic Journalism on Work**

Given the task analysis above, we analyze the impacts of algorithmic journalism on work within a framework consisting of three levels: replacing tasks of journalists, increasing efficiency, and developing new capabilities within the journalism. Algorithmic journalism can take over some tasks. For example, some parts of the tasks T.5, T.11, T.22, T.24, and T.26 that we combined as “generating news reports” may be overtaken by algorithms.

As mentioned, to date, the usage of algorithmic news has been common in reporting about the topics on sports, financial and traffic news [7, 8], whose natures are more appropriate to generate descriptive news reports. Nevertheless, with new more complex algorithms, as emphasized in [5], event-driven narratives, stories, and semantic structures have been included in the generation of algorithmic news. In case preparing algorithmic news as appropriate to blog entries or reviews focusing on literary, musical, or another artwork, through using appropriate data and relevant structures, desired content can be generated, which can be associated with T.26.
Even if algorithmic journalism may generate news content, reviewing news reports is still proper to only journalists. In other words, journalists’ editorial work through interpreting the news generated by algorithms remain the same. Likewise, the tasks T.2, T.4, T.6, and T.14 presented above are still special tasks for human journalists to produce data sources to be used by algorithmic journalists. However, algorithmic journalism may lead to a shift as to the volume of data to be used in generating the news reports. Given that algorithms can use Big Data for deeper investigations [24], journalists may tend to collect a huge amount of information to form data stores consisting of Big Data.

This taking over some tasks of journalists, however, does not correspond to job losses of journalists. In the short term, the concern about unemployment seems unrealistic because these algorithms are used by humans entering data, processing models (i.e., NLG rules or statistical rules, etc.), or checking models; therefore, they are not independent of humans. The Associated Press also noted that “algorithmic journalists” have not caused any human job losses so far. Moreover, senior editors and reporters interviewed in [24] agreed that their work will not be replaced by algorithmic journalism but rather be supplemented by it. The reporters and editors emphasized the lack of “creativity” and “uniqueness” in algorithmic news; thus, algorithms could perform certain manual tasks. They also noted that an editor needs to check the news reports generated by algorithms and to validate their coverage before the publication.

As for increasing efficiency, algorithmic journalism reduces the time and costs by taking over some repetitive and time-consuming work from expensive human staff for the creation of news reports [25]. Moreover, algorithms can analyze a huge amount of data to generate news; therefore, they can produce a lot of news, such as sports news, financial news reports, crime reports. As a result, it is possible to publish stories that otherwise would not have been written. On the other hand, journalists sharing their workload with the algorithmic journalists may have more time for more creative tasks, or for tasks like checking the news generated by algorithms. Thus, the collaboration of human journalists with algorithmic journalists may increase the number of news reports.

As for the quality of news reports, semi-structured interviews conducted with 10 journalists in [25] revealed that there is a concern associated with the lack of creativity and complexity in algorithmic news because this news relies on isolated data, they are repetitive and not interesting. The interviewees emphasized that one of the most important properties of a news report should be using multiple resources [25]. They added that contextual factors cannot be reflected by algorithmic news, thus this news might only present “10 per cent of the story” [25:1247]. The authors of [25] pointed out the lack of human angle in algorithmic news, by expressing that “journalism is about telling stories that involve human beings whose lives are not easily quantifiable into programmable data” [p. 1247].

Jon Bernstein, an editor, writer, and digital media consultant endorses this statement asserting that journalism is not just about presenting information, it is actually explaining what the presented information means [16]. Bernstein adds that automatically generated story content fails to explain the “why” of the story, which requires the ability to analyze and infer. Furthermore, when presented with two articles in an informal poll, NPR listeners preferred the one written by the NPR White House correspondent to the one written by Wordsmith, emphasizing that while Wordsmith is faster than a human journalist, the human-written article was richer and more engaging [23]. Furthermore, some claim that an algorithmic journalist could never beat a human journalist’s style or insight [23]. Thus, even if algorithmic journalism is getting more advanced, it is ambiguous whether algorithmic journalists will be as effective as human journalists.

Even though Wordsmith enables the opportunity users to enter their own data and create their own stories [23], the stories from the input data and the statistical or NLG rules and templates will be monotonous or very similar (because as input data and rules change, the output changes) and lack human creativity, different human emotions, different human perspectives, and different humans’ different writing styles, which make the news move engaging and richer. Although Wordsmith adds emotive languages with appropriate syntax and diction to generate more readable news [16], it still lacks more complex emotions that are specific to humans and vary from person to person.

Finally, algorithmic journalism may bring with it some new capabilities into current journalism practices. For example, if Wordsmith or similar tools to be developed are open to public use, this may lead to a transformation in journalism from news written by journalists in newsrooms and published to everyone to more individual preferences-oriented news. Robbie Allen, the CEO of Automated Insights, indicated that Wordsmith can generate articles much faster than even the fastest writer: it generated more than 1.5 billion pieces of content in 2015, up from 300 million in 2013 [16]. Allen added that instead of writing one story to present it to a million people, Wordsmith provides the opportunity to create individual stories for a million users according to their specific preferences through their participation in the process of news generation [23]. Each story is specific to the user because it is powered by their data [23].

Algorithmic journalism may require journalists to embody new capabilities. Both editors and journalists
interviewed in [24] pointed out the advantages of algorithmic journalism, like it’s difference from purely factual reporting given its availability to use various resources in-depth investigations, which requires human journalists to find these various data sources. Thus, as mentioned before, this requirement may cause a change in journalists’ specific tasks, such as the task concerning data gathering. The volume of data gathered may be increased by algorithmic journalism.

Algorithmic journalism also encourages journalists to improve their technical skills associated with generating algorithmic news. Technical skills, such as the capability of programming and using NLG may be useful to understand the potential of algorithms. Moreover, to benefit from algorithmic journalism technology, journalists should develop themselves to do algorithmic accountability, which is associated with computational thinking, programming, and technical skills required to assess algorithmic decisions [8]. A limited number of journalists have the technical skills to investigate and evaluate algorithms.

Algorithmic journalism also builds new collaborations to produce algorithmic news. For example, a journalist interviewed in [24] highlighted his belief for strengthened future collaborations between reporters and technical staff who helps to execute algorithms for generating news reports. A data journalist interviewed in [24] providing an example regarding how automation can assist the journalists mentioned their collaboration with sports analytics company Opta Sports providing large data sets based on the matches’ scores and parameters related to each player’s performance to produce news reports about soccer matches.

4.3. Ethical Challenges of Algorithmic Journalism

In this section, to answer the second research question, benefiting from the relevant work, we examine the ethical challenges that algorithmic journalism brings with it and possible solutions addressing them. Commonly emphasized relevant ethical challenges detailed discussed in the relevant work section 2.3 were the lack of transparency, privacy, the lack of interlocutors to take responsibility for the violation of ethical rules, problems related to lack of authorship, fake news, news with errors or with mistakes, etc. Also, algorithms may generate biased news because of biased data obtained from biased sources with which the models are fed. In addition, these algorithms may be misused, or accidents may occur while using them. For example, data to be used as input may be tainted with misinformation (false or inaccurate information which unintentionally deceives others), which may result in generating misinformation; input data may not correspond 100% with the templates, or with the NLG rules to be used, which may produce misinformation. Furthermore, input data may be created to manipulate readers according to the data providers’ purposes. Also, these news reports containing manipulated data would be perceived by readers as objective, since they are generated by algorithms, not people. Therefore, this perception might be dangerous in terms of changing peoples’ minds and attitudes towards current events.

Present concerns related to journalistic ethics may be decreased by assigning interlocutors during the generation of algorithmic journalism. For example, the person who inputs the data chooses the algorithm and checks the story may be considered the interlocutor who is responsible for any violations of the ethical rules. After the news is generated and shared, for each part of the content of the news, individuals should apply a verification framework to examine the news, its source, and its context [2].

5. Discussion

In this paper, several adopted models were presented that illustrate the steps that are used to generate news content by algorithms, such as decision trees and NLP, and potential effects on algorithmic journalism on work was discussed. For discussing the impacts on work, we analyzed the tasks defined by O*NET for the reporters and correspondents. Considering the steps for producing algorithmic news and these tasks, we presented a data flow diagram, which can shed light on potential changes in the roles of journalists. For example, as noted in the impacts on work section, some tasks, including writing repetitive news reports in specified fields, such as sports, finance, can be overtaken by algorithms, which can create a human-AI collective intelligence through which humans and algorithms collaborate for generating news reports. Moreover, in addition to journalists, users from the public can create their own news reports by entering their own data by tools, such as Wordsmith, which can cause shifts in the traditional journalism practice, from more general to more individual-oriented news reports.

Certain tasks, such as, “T.21. Communicate with readers, viewers, advertisers, or the general public via mail, email, or telephone;” however, are still classified as journalists’ roles. In [24:8], this role is emphasized by referring to journalist interviewees saying that “with data research, what follows is always traditional research. I need to speak to people and let those involved have a say. That’s the first bit. And what comes on top of that are technical skills,” and “without the willingness to meet new people and to approach them directly, [and] without curiosity, nothing is possible.” Schapals and
Porlezza [24:3] connected these roles with journalists’ normative roles (defined as “what journalists should do and what society expects of them”), such as meeting with new people and cognitive roles (defined as “what journalists want to do, and how this idealized scenario corresponds to the normative roles expected of them”), such as curiosity, both contribute to the proper workings of a democratic news production environment [24]. Additionally, in [24], this role is connected with the journalists’ public service role, meaning the role of acting as servants for the public and as ‘watchdogs’ over society. The common expressed belief of journalists interviewed in [24] was in support of algorithmic journalism: due to the ability of accessing huge amounts of data which can help offer in-depth investigations, the roles of serving to the public and as ‘watchdogs’ over society might not only remain active, but may actually be strengthened.

It is controversial whether the use of algorithmic journalism overall is beneficial or not. Whereas some have advocated that algorithmic journalists will augment journalists by helping them to generate news at higher speed using Big Data. The study [24] based on the interviews reflecting the German newsrooms’ attitudes towards algorithmic news showed that journalists believe that algorithmic journalism will advance their roles through supplementing them. The interviewees believe that they benefit from automation in a method that protects their authorship and interpretable role; namely, journalists control the algorithms that may be used in Big Data projects and edit the work performed by algorithms.

Ref [24] emphasizes also that the supplementing role of algorithms may depend on different circumstances. For example, a journalist suggested to use algorithmic journalism to publish on smaller, local newspapers, on the other hand, for publishing on larger organizations, he suggests using semi-automated news in which content is personalized by journalists [24]. One of the journalists from Zeit Online interviewed in [24] suggests that algorithms could be used as automatic alerts informing journalists about sudden events (e.g., earthquakes), thus supplementing them.

Arguments associated with the negative effects of the news generated by algorithms also exist. Some believe that algorithmic journalism will not be as effective as news written by human journalists because algorithmic news does not go beyond the descriptive coverage [24] and does not reflect emotions, values, creativity, and so forth, thus lacking of human angle and real storytelling. As a result of this, the quality of the news produced by the algorithms may be viewed as insufficient.

Moreover, other problems with algorithmic journalism concern authorship, credibility, quality, unemployment that affects journalists, and ethical concerns, such as risks of violation of journalistic ethics, the lack of interlocutors to take responsibility for the violation of ethical rules, fake news, news with errors or with mistakes, etc. However, these potential problems are not unsolvable. In the example of Wordsmith, if a user enters data and creates a story, in this case, the user may be considered the author of the story. Also, as emphasized before, based on the common belief of journalists interviewed in [24], the authority will still belong to journalists checking the algorithmic news, providing the data to algorithms and reviewing the algorithmic news; hence, these journalists can be considered as authors of this news, thus the concern regarding authorship may be eliminated. The quality and credibility of the stories may depend on the data and data models.

In sum, if algorithmic journalism technology is used under the control of journalists checking results and following ethical rules, we can envision that this technology can be useful for creating a huge amount of content at higher speed with fewer costs. Thus, as a practical implication, we suggest journalists to determine general ethical rules internationally accepted that address the journalists’ common concerns and to obey these rules.

As a limitation of this study, we admit that this study’s findings were not constituted based on the direct empirical data. Thus, as a research implication, we recommend researchers to conduct more comprehensive empirical studies (e.g., surveys or interviews) in the future to explore the impacts of algorithmic journalism on specific tasks of journalism.

6. Conclusion

For years, journalists have been benefitting from many technological tools, such as bullhorns, tablets, computers, cameras, and phones, to gather, produce, present, and distribute information to the public [26]. Current presentational formats, including hooks, listicles, gifs, podcasts, virtual and augmented reality, conversational interfaces, and data visualization, are utilized to produce more attractive news [26]. Today, more advanced technologies, such as NLG (natural language generation) based on AI, are used to generate news content. Moreover, because of ongoing improvements in AI – thanks to Big Data, advanced algorithms, and more powerful computers – new programs may emerge that can enhance algorithmic journalism [13].

Algorithmic journalism may be spreading out more because of its speed and its power to deal with huge amounts of data, which provide deeper, more specific, and immediately available information, which can benefit society, as long as ethical rules are followed and necessary measures are taken, such as checking input,
output, and models regularly to eliminate ethical concerns, such as violations of transparency, verification, privacy, bias, etc. Moreover, automation may call for human skills, such as judgment, curiosity, and skepticism, so that we can continue to access succinct, comprehensive, and accurate news. This paper can contribute to offering an initial understanding of how algorithmic news is created, how algorithmic journalism technology impacts journalists’ work, and what type of ethical issues brings with it and their possible solutions.

7. References


[23] Reynolds, E., “Wordsmith’s ‘robot journalist’ has been unleashed”, 2015, pp. 3.


[27] “27-3022.00 - Reporters and Correspondents”, https://www.onetonline.org/link/details/27-3022.00