Review

Algorithmic Journalism—Current Applications and Future Perspectives

Efthimis Kotenidis [ID] and Andreas Veglis *[ID]

Media Informatics Lab, School of Journalism & Mass Communication, Aristotle University of Thessaloniki, 541 24 Thessaloniki, Greece; kotenides@jour.auth.gr
* Correspondence: veglis@jour.auth.gr

Abstract: Journalism, more so than other professions, is entangled with technology in a unique and profoundly impactful way. In this context, the technological developments of the past decades have fundamentally impacted the journalistic profession in more ways than one, opening up new possibilities and simultaneously creating a number of concerns for people working in the media industry. The changes that were brought about by the rise of automation and algorithmic technology can mainly be observed in four distinct fields of application within journalism: automated content production, data mining, news dissemination and content optimization. This article focuses on algorithmic journalism and aims to highlight the ways that algorithmic technology is being utilized within those fields, as well as pointing out the ways in which these developments have altered the way journalism is being exercised in the modern world. The study also discusses challenges related to these technologies that are yet to be addressed, as well as potential future implementations related to algorithmic journalism that have the capacity to improve on the foundation of automation in the news industry.

Keywords: algorithmic journalism; automated content production; natural language generation; data mining; news dissemination; content optimization

1. Introduction

Journalism is a profession that has always been shaped by technology throughout history (Pavlik 2000). Despite its constant and very close relationship to technological advancements, however, the past decade has seen an especially large shift in the field, with many of the core elements of the journalistic profession being redefined (Deuze and Witschge 2018). The introduction of new and innovative technologies such as artificial intelligence and natural language generation (NLG) was partially responsible for this transformation. These advancements have brought very noticeable changes in the way the journalistic profession is being exercised, particularly because of their influence in news production, as well as news dissemination (Dörr 2015), by creating numerous new opportunities when it comes to gathering and consuming news (Spyridou et al. 2013). Historically speaking, computerization and the rise of automation has shown us that technology is prone to taking over routine tasks (Frey and Osborne 2017) and the same has proven to be true when it comes to journalism as well. Over the years, as artificial intelligence started to improve and evolve, various automated algorithms begun to substitute human workers in the field, by taking over different tasks (Graefe 2016). These tasks varied in complexity over the years, starting from more streamlined processes such as collecting basic information, and moving into more demanding duties such as completely constructing news stories from scratch with modern algorithms, to the point where nowadays each different step of the news production process can be replicated by a machine (Van Dalen 2012). All of the procedures described above can be encompassed within the term "algorithmic journalism", which is going to be the focal point of this study.

The main focus of this paper is to thoroughly examine all aspects of algorithmic journalism in order to identify its main areas of application. Simultaneously, it aims to analyze algorithmic usage in the field and underline exactly how it has impacted journalism, as well as what potential future implementations of algorithmic tools might mean for the field as a whole. For the needs of the study, we have employed a literature review (Grant and Booth 2009). The review includes a significant number of scientific articles that have been published in peer-review journals and scientific conferences. The selection of the articles was based on the existence of specific keywords in the titles and/or abstracts of manuscripts. Through the process of the literature review, this study attempts to identify the utilization of algorithmic technology in the media sector for summation and for distinguishing potential gaps. Thematic analysis has been employed in the presentation of the findings. Overall, seventy articles were examined, covering a time period from 1970 to 2020, with 51 of them originating from scientific journals and 10 of them being part of conference proceedings.

In order to achieve the above, this paper is organized in five main sections, including this introductory chapter. The following section is going to focus on the definition of algorithmic journalism and what exactly the term means within the boundaries of the journalistic profession. Section 3 provides a comprehensive analysis of the four main areas where algorithmic technology is more commonly applied in journalism, whereas Section 4 focuses on the challenges these technologies face at the present time, as well as potential future implementations of them. Finally, Section 5 closes the article, by summarizing the points that were brought up earlier in the study.

2. Definition of Algorithmic Journalism

Algorithmic journalism is a term that attempts to describe the procedures that have been brought about by recent technological changes in the field of journalism. Some researchers such as Graefe (2016) define algorithmic journalism as "the process of using software or algorithms to automatically generate news stories without human intervention", not accounting for the original programming of the software of course. While definitions such as this one aptly describe what is perhaps the most important aspect of algorithmic journalism—that of automated content production—they leave out some of the other applications of such technologies in the field. For this reason, the term algorithmic journalism is usually interchangeable with a variety of similar terms found in related literature, such as computational journalism, robotic journalism and automated journalism (Anderson 2012). In an attempt to describe their wider scope of applications, a more generalized and inclusive definition for these terms would be "the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism" (Hamilton and Turner 2009). Definitions of this sort encompass the large variety of different technological applications one can encounter in the field of journalism today, while at the same time acknowledging the valuable contribution of human workers in these procedures.

While broadening the definition of such a term can prove invaluable in understanding all of its facets, it should also be noted that an extremely broad definition should be avoided, since it does not help in narrowing down the exact focus of the subject matter. For example, ever since the dawn of digital technology, the term computer-assisted reporting (CAR) was used to describe any sort of digital assistance journalists utilized in their workflow, including the use of personal computers for simple tasks such as online research (Garrison 1998). That is the reason researchers such as Diakopoulos (2011) attempt to draw a line between the two terms, highlighting the fact that algorithmic journalism—while still being inclusive of the term CAR—focuses more on the processing capabilities of modern software, as opposed to the more mundane facets of technology utilization from journalists, such as storing and accessing data.

While computational technology has always been a valuable tool for the professional journalist, and in the past 10 years, it has helped tremendously with productivity in

the field of journalistic work (Lindén et al. 2019); the relatively new phenomenon of the complete automation of the news production process has created a lot of heated debates among journalists and researchers alike. The division of labor has seen a major shift, as algorithms are becoming more and more capable of executing tasks that were once the sole responsibility of human workers and the implications of this development have led many practitioners to question whether a future where newsroom jobs are entirely automated is a good thing (Graefe 2016). In order to better understand the roots of this debate, as well as to evaluate the ways in which journalism has changed over the past decade, it is important to examine the individual areas in which algorithms and automation have been the most impactful, within the confines of the journalistic profession.

3. Areas of Application

Journalism has changed vastly over the past years and the responsibility for this change hinges mostly on the very significant impact modern technology has had in the news industry. What follows in an analysis and review of the main areas in which computational technology has brought the most notable changes in the field (Figure 1). It should be noted that algorithmic usage in journalism can potentially go far and beyond the categories that are listed here. The following, however, are the ones that appear to have the most relevant services in regard to the modern news industry (Lindén 2017). Specifically, those are:

1. Automated content production;
2. Data mining;
3. News dissemination;
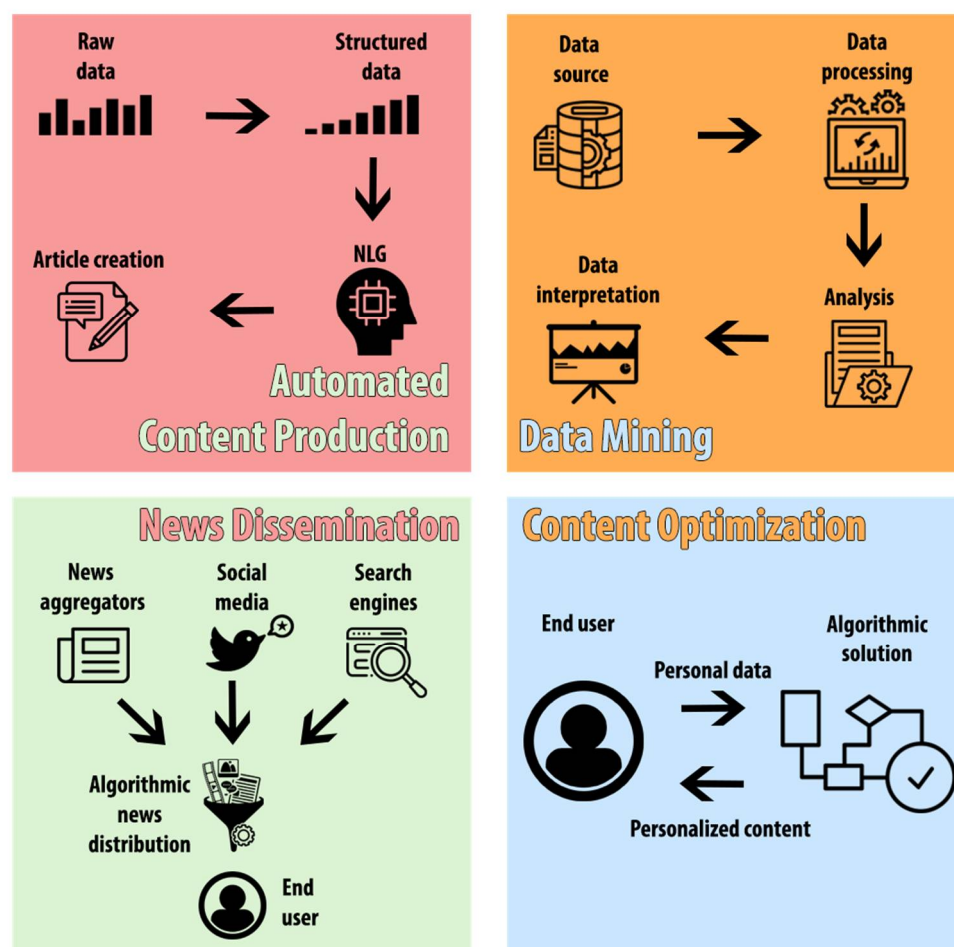4. Content optimization.

Figure 1. Fields of application for algorithmic journalism.

3.1. Automated Content Production

The automation of the news creation process is perhaps the most important—and as a result, the most controversial—of all the fields of application for algorithmic technology in journalism (Montal and Reich 2017; Schapals and Porlezza 2020). In the grand scheme of things, this particular field of application is considered a relatively recent development in the field of journalism (Ali and Hassoun 2019; Graefe 2016) and it consists mainly of algorithms and automated software that are capable of creating news stories on their own (Diakopoulos 2019).

One of the most well-known examples of early applications for automatic content production is that of "Quakebot", a program that was created on behalf of the Los Angeles Times in 2014. Its purpose was to closely monitor data from the US Geological Survey in an attempt to identify instances of seismic activity and proceed to write and publish simple reports on them (Otter 2017). Since then, automatic content production has taken major steps forward, to the point where some of the biggest contributors to the industry such as Forbes and The New York Times often rely on algorithmic production for their content, with the end result being almost impossible to distinguish from human writing (Clerwall 2014).

The basis for the innovations in automated content production is a technology called "natural language generation" or NLG for short. Natural language generation is defined as "the automatic creation of text from digital structured data" (Caswell and Dörr 2018) and it is a technology that first made its appearance in the 1950s within the context of machine translation (Reiter 2010). NLG has seen exponential growth in the past few years and in light of these developments, many industries have begun to utilize it alongside artificial intelligence to further improve their products and services, with the news media industry being no exception to this rule (Diakopoulos 2019).

The adoption of these technologies by the journalistic profession brought with it a number of advantages, including a significant increase in productivity thanks to the publishing of stories without any human intervention (Ali and Hassoun 2019) as well as the ability to allow journalists to redefine their core skill set (Van Dalen 2012) and provide them with more creative freedom in their work (Milosavljević and Vobič 2019), since computers were able to execute part of their responsibilities by taking over routine tasks (Glahn 1970). Those advantages also seem to coincide with the increasingly high market demands for fast and accurate news stories, making algorithmic news production even more beneficial (Clerwall 2014; Diakopoulos 2019).

Thanks to the above, algorithmically generated news started to become a near necessity in the modern news production cycle (Zangana 2017), which, in turn, has led to various forms of controversy from members of the news industry. The main discussion point between journalists and people that are employed in the news industry as a whole is the possibility that the automatization process might render human workers in the field obsolete (Veglis and Maniou 2019). There have been many arguments recorded in related literature when it comes to this topic, and many workers have also voiced their opinion, suggesting that the increasingly dominant role of algorithms in the newsroom will pose a serious threat to the future of human journalists (Kirley 2016). On the opposite end of the spectrum, a number of researchers seem to suggest that those fears are mostly unfounded, pointing out that artificial intelligence and algorithms are only going to enhance journalistic practice in the long run instead of replacing it (Hansen et al. 2017).

Drawing a line between what might be a useful innovation and what might pose a threat to the industry due to the potential loss of jobs is certainly no easy task, and that is perhaps the reason behind this apparent split in the existing literature, with many researchers pointing out the benefits of automation, and others focusing on the potential danger it encompasses for the employees of the media industry. It is certain that automated content production plays a major role in the news production process nowadays, and it is commonly agreed by researchers that automation will hold a critical role in the future of news agencies (Liu et al. 2017). As competition within the industry continues to rise,

the only way to keep up with the ever-increasing demand for more news stories seems to be the utilization of automated content production technologies. The question remains, however, as to how the industry is going to adapt to these new conditions of automation, as the displacement of employees and an overall reduction of the workforce is indeed inevitable based on current projections (Carlson 2015), as machines become more and more capable in substituting human workers in specific tasks.

There are a number of views shared by researchers and employees in the media industry that tend to challenge the arguments presented above, regarding the ability of algorithms to "free" journalists and allow them more time to pursue more investigative tasks (Schapals and Porlezza 2020). These concerns mostly stem from the fact that computational technology is shaping journalism into a more streamline and sterile process, one that does not necessarily require human input in order to function, and they bring up some very valid points regarding the skill set that a modern journalist is expected to have in order to compete in this environment. Taking that into consideration, the fact that automation will make a number of jobs obsolete given enough time seems to be an inevitable outcome. While the way the industry intents to deal with this problem still remains to be seen, perhaps one potential solution to it lies in the adjustment of expectations and the redefining of the term "journalistic labor". As Carlson (2015) puts it, "Automated journalism requires the transformation of journalistic labor to include such new positions as "meta-writer" or "metajournalist" to facilitate automated stories". This point of view suggests that in order to achieve a fully symbiotic relationship between human workers and machines, a middle ground has to be reached, specifically one where media industry workers need to reevaluate their priorities and develop a skill set that supplements algorithmic news production, instead of attempting to compete head-on with it. In accordance to what Van Dalen (2012) has stated, this can be seen as an opportunity for workers to redefine their core skills and work in tandem with algorithms, as ultimately, these programs are fundamentally different from humans, since they lack traits such as creativity, flexibility and analytical thinking, which would mean that in order to achieve the best and overall most efficient result, both parties would need to work together and cooperate.

The fact that these programs lack traits such as creativity, flexibility and analytical thinking is an important factor that separates them from humans (Van Dalen 2012); as such, these technologies do not present an immediate threat to the practitioners of the journalistic profession (Ali and Hassoun 2019).

Despite how important automated content creation has been for the industry, it is apparent that algorithmic journalism is not limited just to the creation of automated news stories (Jamil 2020). There are other important fields of application for these technological innovations that that have also impacted journalism in a major way, which will be examined below.
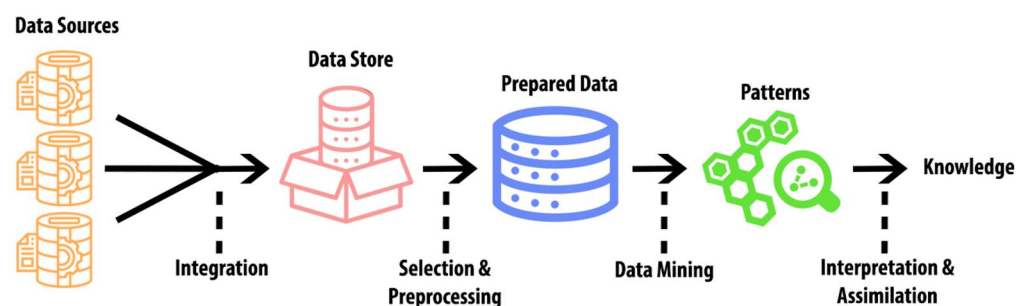
3.2. Data Mining

One of the most defining characteristics of the information age that we are currently undergoing is the so-called "data explosion", which refers to the constant increase of widely available data on the internet, with some sources approximating that the digital universe roughly doubles in size every 18 months (Zhu et al. 2009). Data, however, should not be mistaken for information (Aljazairi 2016). Within this ever-increasing landscape of available resources, journalists are struggling more than ever to separate clutter from actually useful information (Chen and Liu 2004), and this is where the need for procedures such as data mining starts to become apparent.

According to Bramer (2007), data mining is a central part of a broader process called "knowledge discovery" and it refers to the extraction of useful information from a larger subset of data (Figure 2). There are many applications for this type of technology in journalism, with the most obvious one being the acquisition of specific information from large databases. The case of "Quakebot" that was mentioned above also constitutes a very good example of data mining, despite the fact that it is mostly known to be an

instance of automated content production, since the program was able to single out and use information form a much larger dataset (which was all of the data provided by the US geological survey). Chatbots and other similar automated agents have been utilized extensively in these procedures (Veglis and Kotenidis 2020).

Figure 2. The knowledge discovery process according to Bramer (2007).

Other than this more obvious use case, however, the technology behind data mining can also be utilized for various other complex tasks related to journalism. For example, there are instances where datasets are too massive for humans to even comprehend, because of characteristics such as their volume (terabytes–petabytes) or their velocity (being created in real time), and this makes algorithmic data mining the only reasonable way to tackle these so-called "Big Data" (Kitchin 2014). Journalists often find themselves working with these types of data sets as part of their job and data mining can help them uncover previously unseen connections between variables with high statistical significance, which in turn can allow them to test complex ideas and hypotheses (Latar 2015).Data mining also has the ability to enable other fields of application found in algorithmic journalism since it can be used to discover new social trends and automatically target specific consumers who might find the content more relevant (Latar 2015), as well as being used in conjunction with automated content production, as seen in the example presented earlier in the manuscript.

While procedures such as data mining have mostly been recognized as strictly beneficial to the journalistic cause, there is still a discussion to be made regarding their ethical side. As Kennedy and Moss (2015) point out, the undoubtable usefulness of algorithmic mining—specifically in online spaces with user interactivity such as social media—can occasionally be overshadowed by privacy considerations regarding user surveillance that could lead to social discrimination. Metadata analyzed in this way can sometimes be even more valuable than the content that is being shared. Of course, as is the case with any tool, the intent behind the usage of data mining software is equally as important as any practical concerns surrounding it and that is the reason that studies such as the one mentioned above propose the democratization of these procedures via the introduction of regulations and more meticulous public supervision.

In addition to the above, the question of accessibility that has been raised earlier in the manuscript also applies to these advanced tools. Similarly to algorithmic news production, the introduction of Big Data and the appropriate procedures required to analyze them has also impacted the news industry in a big way, not only in the productivity department, but also in the skills required to work in this new and rapidly changing environment (Hammond 2017). In order to be able to understand the complex information hidden in large datasets, workers in the news industry should be able to utilize modern tools and special software that will allow them to take full advantage of Big Data in order to supplement their reporting and information-gathering procedures (Veglis and Maniou 2018). This argument is closely related to the considerations that surround automated content production, in the senses that the evolving media landscape is going to require workers to acquire a much more specialized role in order to stay competitive in this increasingly automated work environment. Much of what has been said about automated programs replacing human workers in the case of content production can also be said here, although

in the case of algorithmic data mining, there are some notable exceptions such as the analysis of Big Data itself. In these instances, software agents seem to only expand the capabilities of the modern journalist, without any risk of replacing actual workers, since Big Data and other similar concepts are by their very nature unable to be processed by humans and would otherwise be inaccessible without the help of algorithms (Kitchin 2014).

3.3. News Dissemination

In this day and age, the internet accounts for a very large portion of daily media consumption (Gaskins and Jerit 2012) and as such the way the dissemination of news is handled proves to be exceedingly important (Orellana-Rodriguez and Keane 2018). There are three main platforms through which the majority of internet users receivetheir news content from, namely: news aggregators, search engines and social media sites (Foster 2012). These digital intermediaries all have something in common: they largely rely on algorithms and automated systems in order to appropriately distribute content to their users (Cádima 2018).

As media companies started to shift their focus on online news and the implementation of more interactive features (Deuze 2005), these automatic news dissemination technologies proved to be a major driving force for journalism since news organizations started to utilize them more and more (Carlson 2018). The advantages that emerged in the field of journalism through the use of these innovations became apparent quite quickly. Specifically, news outlets were able to utilize algorithms in order to automatically and systematically disseminate news on social media and other similar platforms, by using software agents called "news bots". These programs are capable of distributing news and information to a large audience, as well as interacting with users in various ways and ensuring high visibility for the content in question, thereby supplementing the news dissemination process and helping media agencies to reach as wide an audience as possible (Lokot and Diakopoulos 2016).

Controversy has also been observed in this field of application, although perhaps not to the extent of automated content production. Specifically, concerns have arisen from researchers over the years regarding the role of algorithmic news distribution technology as a "gatekeeper" of news (Nechushtai and Lewis 2019; Cádima 2018), the accountability and the impartiality of these programs (Diakopoulos 2015) as well as ethical considerations regarding algorithmic transparency (Diakopoulos and Koliska 2017) and the role these agents play in the spread of fake news and misinformation (Shao et al. 2017; Shin and Valente 2020; Fernandez and Alani 2018).All of the above constitutes well-funded criticism related news dissemination that has yet to be addressed in a meaningful way. When it comes to news gatekeeping in particular, Cádima (2018) brings up an important point regarding the intermediation issue. As digital intermediaries are estimated to be redirecting more than 70% of internet news traffic, it is difficult to ensure that news circulation will remain democratic going forward. This poses a lot of questions about the future of journalism that are related both to quality deterioration, as well as censorship issues that could potentially affect a very large subset of the population. Ensuring that communication channels remain open and not allowing any third parties to consistently prioritize certain voices over others will prove vital for the future of the journalistic profession. Ultimately, however, an agreed-upon standard for humans as news gatekeepers does not exist, and this fact makes it all the more challenging to assess the performance of algorithms in this regard (Nechushtai and Lewis 2019).

3.4. Content Optimization

Personalized content for individual recipients is not a new idea in the media industry, as some researchers have suggested functioning models for it even before the turn of the 21st century (Bharat et al. 1998; Billsus and Pazzani 1999). Despite this, however, it was not until the past few years that developments in algorithmic technology allowed news providers to target specific audiences on a large scale and deliver customized news experiences for

them, thanks to the internet's ability to provide almost real-time recommendations and information from all over the world (Li et al. 2011). These personalized news content services have proved to be very useful because they can save time for the end used by drastically reducing the amount of irrelevant information and provide content only for subjects that are of interest (Jokela et al. 2001).

Content optimization for users usually works in a similar manner to search engines, which utilize automated ranking algorithms in order to return the most relevant results for a user's search. Using a similar structure, personalized news content and online ads are served to specific users with the use of automated algorithms (Agarwal et al. 2008). Content optimization with the help of algorithmic technology has also been observed in other parts of the news production process, as some organizations utilize algorithms for tasks such as A/B testing for article headlines in order to better gauge their effectiveness (Lokot and Diakopoulos 2016). The prime use for this technology, however, has been the delivery of personalized news content through customized newsfeeds or automated agents such as chatbots. These automated bots in particular have proven to be very effective in engaging with audiences by providing more interactive and personalized instances of news and articles as opposed to the traditional methods of content consumption (Jones and Jones 2019).

Even though this technology provides a user-friendly way of consuming more relevant content, there have been a number of concerns regarding its use that are worth addressing. First off, some privacy concerns have been brought to light by users over the years in regard to content optimization. Specifically, those concerns are related to the way these algorithmic solutions function, since most content optimization systems from media organizations and other companies alike rely on the collection of personal data in order to fulfill their duties (Das et al. 2007). Furthermore, the personalization employed by these algorithms often remains unnoticed by the users (Powers 2017), which further feeds into this issue. That is the reason many researchers, such as Diakopoulos and Koliska (2017) and Graefe (2016), have started to advocate for algorithmic transparency over the past years, since many users do not feel comfortable with the idea of "being watched" by automated programs without them being notified while they are browsing the internet, even if that action ultimately aims to benefit them with more streamlined recommendations.

The privacy concerns mentioned above are likely to grow in scale with each passing year as technology gradually envelops more and more aspects of daily life, and as such, it is important for algorithmic transparency to be established as one of the pillars upon which future innovations can be developed, in order to avoid further frictions. Despite their importance, however, these concerns are not the only ones that were brought to light when it comes to personalization algorithms. Another relevant issue in this field of application has to do with the content that is being distributed. Specifically, researchers have noted that the constant stream of personalized content has the potential to negatively affect the news ecosystem, since it has been known to reduce news diversity for recipients and consequently lead to partial information blindness (Haim et al. 2018). This phenomenon became widely known with the term "filter bubbles", with similar theoretical constructs such as "news echo chambers" describing constant user exposure to like-minded opinions (Garrett 2009). These online environments that stand devoid of varied viewpoints constitute a serious criticism regarding news personalization, since they tend to reinforce the user's opinion on specific matters, and usually offer no counterpoints, or even alternative viewpoints to the one they have chosen to adopt. Even though this phenomenon is not exclusive to these technologies, or even to the internet as it can be observed in other media as well, the nature of online personalized content delivery seems to be enhancing this particular problem. To put it in simpler terms, while algorithmic personalization caters to the needs of the user and creates a more enjoyable and customizable experience, it also simultaneously encompasses them in their own "bubble" and prevents them from challenging their beliefs. This criticism puts the model of personalized news delivery into question, as it can be the

epicenter of some serious ramifications in the future that can range from the spread of misinformation to the potential fragmentation of the public opinion (Graefe 2016).

4. Challenges and Potential Future Implementations

Even though algorithmic technology has come a very long way over the past years, there are still a number of challenges it needs to overcome. The majority of them are related to automated content production since that field of application is—by its very nature—very complicated and demanding. One of the main limitations of automated journalism is its dependence on structured data. Modern algorithmic solutions rely heavily on structured information in order to compose articles, and because of this, reliance topics cannot be covered unless adequate structured data exist for them (Graefe 2016). Similarly to data availability, data quality is also very crucial for this procedure, since poor data quality will likely result in less accurate reporting (Leppänen et al. 2017). Moreover, while the algorithmic ability to mimic human writing and produce journalistic content has steadily developed over the years, there are still a few key areas in which these programs are still falling behind compared to humans, and according to researchers, this fact is not expected to change soon. These areas are mainly analytical thinking, flexibility and creativity (Van Dalen 2012), as well as the ability to draw conclusions or ask questions and explain new phenomena based on the provided data (Graefe 2016). Despite how efficient automated journalism can be, especially in events that have a repetitive or predictable structure such as sports games or weather reports (Graefe and Bohlken 2020), these facts still create a separation between news-writing algorithms and humans at the present time. On top of that, there are also editorial challenges related to automated content production that could prove to be even more difficult to overcome than the technical limitations (Caswell and Dörr 2018). In order to move past these problems, journalists are expected to develop the required "computational thinking" in order to accommodate for any shortcomings the algorithms may present in that department and work hand-in-hand with these programs in order to ensure the best possible results.
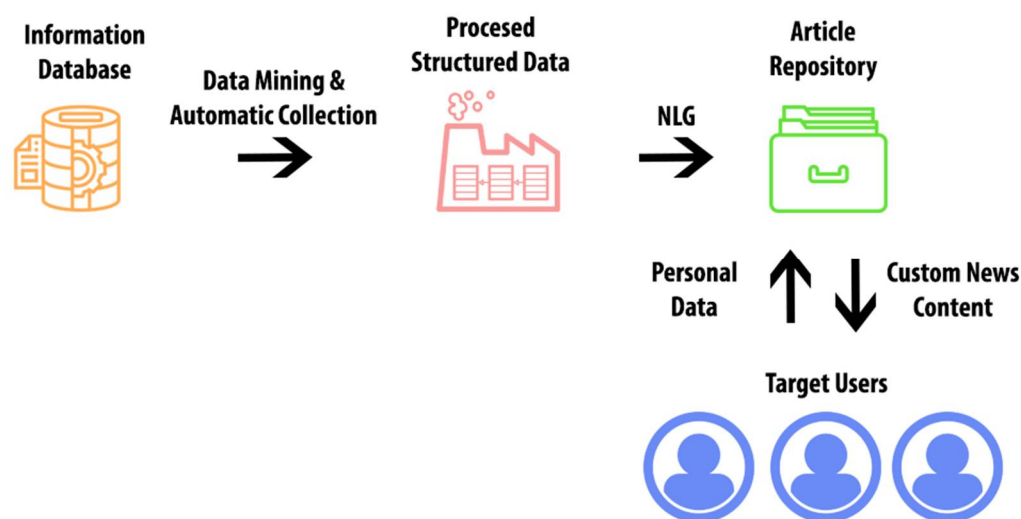
Other than automated content production, challenges and limitations also exist in regard to other fields of application as well. When it comes to data mining, while algorithms excel at discovering connections between multiple variables, oftentimes, the results they offer can be meaningless, or even lead to the wrong conclusions altogether (Latar 2015). The reasons behind these false discoveries can vary from wrong questions to incorrect data, or artificial intelligence procedures. This highlights the fact that, no matter how capable these tools may be, correct utilization of them remains paramount and often requires adequate knowledge on behalf of the media workers in order to achieve the expected results in the context of journalistic research.

Regarding news dissemination and content customization, there have been concerns by researchers regarding automated news. Specifically, the advent of personalized content has the potential to lead to fragmentation of the public opinion (Graefe 2016). The reason behind this is the possibility for the creation of "news echo chambers", which are online environments devoid of other viewpoints, except the ones that the recipient of the news content agrees with (Garrett 2009). Even though this phenomenon is not exclusive to these technologies, or even to the internet, as it can be observed in other media as well, the nature of online personalized content delivery seems to be enhancing this particular problem.

Despite all of the above, algorithmic technology remains a very promising field when it comes to the evolution of the profession and researchers believe that artificial intelligence and automation will help journalists overcome some of the fundamental problems contemporary journalism is faced with, such as the overabundance of information and the related credibility issues (Ali and Hassoun 2019). Even though the introduction of more sophisticated news algorithms down the line is bound to cause some turbulence in the industry, in a similar way to almost all other professional fields impacted by automation (Ford 2015), it is widely believed that the potential that these programs carry along with them will help journalists to produce news at a quicker pace, on a larger scale and with

fewer errors overall (Lewis et al. 2019). Researchers have noted that as the technology develops, algorithms can be utilized in order to cover events that would be uneconomical to cover under the current circumstances (such as specific sports with low attendance or interest), as well as to create audiovisual reports in addition to text (Thurman et al. 2017). This future development has the potential to greatly expand the news writing scene, as autonomous programs could be tasked with "writing stories in spaces where no one is writing stories" by turning raw data into compelling narratives (Carlson 2015). While these developments are no doubt beneficial from the perspective of media pluralism, this surge in news production, in combination with some of the concerns mentioned earlier in the article regarding content dissemination and optimization, could potentially accumulate into another sort of issue altogether. Specifically, since news availability will increase exponentially as time passes due to faster production and better delivery methods, there exists a real risk of information overload in the media landscape. As stated by Graefe (2016), "Automated journalism will substantially increase the amount of available news, which will further increase people's burden to find content that is most relevant to them". This angle is certainly an important one to consider going forward, especially because the advent of fake news and misinformation is likely to further compound the problem.

Finally, when it comes to future implementations, another venture that can potentially prove to be very promising is the creation of a fully autonomous news system. Such a system would be able to combine different areas of application such as data mining, algorithmic content production and news dissemination and optimization in order to sort through information, write reports based on the collected data and distribute the final product to appropriate audiences, all without the need for human intervention (Figure 3). While an implementation such as this one will certainly prove to be challenging, examples such as the one presented with "Quakebot" prove that such a concept can work in theory, especially when it comes to events with a repetitive structure, such as weather reports (Graefe and Bohlken 2020).

Figure 3. Structural example of a fully autonomous news system.

5. Conclusions

Technology and automation have had a tremendous impact on all aspects of human life. The turmoil caused by these developments has impacted nearly every industry, and journalism is no exception to this rule (Ford 2015). Within these circumstances, four distinct fields of application for algorithmic technology emerged in regard to journalism. This paper has attempted to explore these areas of application by highlighting their effects on the workflow of modern journalists, as well as the changes they have brought to the industry in general.

Automated content production is a revolutionary, albeit controversial, development that sees algorithms mimicking human writers and creating news stories based on structured data. This technology has led to a number of improvements for the industry, with media organizations relying on such programs in order to flesh out their news schedule and allow journalists to pursue more investigative tasks (Hong and Oh 2020; Jung et al. 2017). At the same time, however, the nature of this type of automation has brought up concerns regarding the potential substitution of human workers in favor of automated algorithms.

Data mining is a technology that is becoming more and more relevant within the context of journalism, as datasets become bigger and bigger with each passing year. This technique allows for the analysis of large amounts of data with the purpose of isolating and extracting useful information from them. Workers in the media field have found data mining to be an invaluable tool that allows them to tackle more complex problems by uncovering hidden parameters, as well as work with concepts such as "Big Data", which are datasets that would otherwise prove to be incomprehensible for humans because of their massive scope.

News dissemination is yet another field where algorithmic technology has taken over the traditional journalistic procedures. News aggregators, search engines and social media sites all employ algorithmic technologies in order to ensure better distribution as well as higher visibility for the content shared through them. At the same time, news bots and various other automated agents are employed in social media with similar goals. All of these changes act as a segue to the fourth and final field of application that was explored in this paper: content optimization. Different users have different preferences in regard to the content they consume; as such, algorithmic technology constitutes a perfect fit for the situation. Newsfeed customizations as well as personalized content delivery have provided a more engaging experience for audiences while at the same time ensuring high visibility for relevant content.

While the technologies mentioned above have already revolutionized the way the journalistic profession is being exercised, there are still a number of obstacles that need to be overcome, not all of which are technological in nature. Except from the obvious technical limitations, a lot of editorial as well as ethical considerations have made their appearance, signifying that the landscape of automation is simultaneously very promising and very challenging as well. As skepticism surrounding privacy concerns and journalistic labor reaches an all-time high, it is important for algorithms to remain as transparent and well-regulated as possible, in order to continue their development in harmony with traditional journalistic values and ultimately fulfill their potential by helping journalists to overcome some of the fundamental limitations of the profession and advance journalistic work beyond what is currently possible.

References

Agarwal, Deepak, Bee-Chung Chen, Pradheep Elango, Nitin Motgi, Seung-Taek Park, Raghu Ramakrishnan, Scott Roy, and Joe Zachariah. 2008. Online models for content optimization. In *Advances in Neural Information Processing Systems*. Vancouver: NeurIPS, pp. 17–24.

Ali, Waleed, and Mohamed Hassoun. 2019. Artificial intelligence and automated journalism: Contemporary challenges and new opportunities. *International Journal of Media, Journalism and Mass Communication* 5: 40–49. [CrossRef]

Aljazairi, Sena. 2016. Robot Journalism: Threat or an Opportunity. Master's thesis, School of Humanities, Education and Social Sciences, Örebro University, Örebro, Sweden. Available online: https://www.diva-portal.org/smash/get/diva2:938024/FULLTEXT01.pdf (accessed on 18 February 2021).

Anderson, C. W. 2012. Towards a sociology of computational and algorithmic journalism. *New Media & Society* 15: 1005–21. [CrossRef]

Bharat, Krishna, Tomonari Kamba, and Michael Albers. 1998. Personalized, interactive news on the web. *Multimedia Systems* 6: 349–58. [CrossRef]

Billsus, Daniel, and Michael J. Pazzani. 1999. A hybrid user model for news story classification. In *Um99 User Modeling*. Vienna: Springer, pp. 99–108. [CrossRef]

Bramer, Max. 2007. *Principles of Data Mining*, 1st ed. London: Springer, vol. 180, p. 2.

Cádima, Francisco Rui. 2018. Journalism at the crossroads of the algorithmic turn. *Media & Jornalismo* 18: 171–85. [CrossRef]

Carlson, Matt. 2015. The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism* 3: 416–31. [CrossRef]

Carlson, Matt. 2018. Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism. *New Media & Society* 20: 1755–72. [CrossRef]

Caswell, David, and Konstantin Dörr. 2018. Automated Journalism 2.0: Event-driven narratives: From simple descriptions to real stories. *Journalism Practice* 12: 477–96. [CrossRef]

Chen, Sherry. Y., and Xiaohui Liu. 2004. The contribution of data mining to information science. *Journal of Information Science* 30: 550–58. [CrossRef]

Clerwall, Christer. 2014. Enter the robot journalist: Users' perceptions of automated content. *Journalism Practice* 8: 519–31. [CrossRef]

Das, Abhinandan S., Mayur Datar, Ashutosh Garg, and Shyam Rajaram. 2007. Google news personalization: Scalable online collaborative filtering. Paper presented at the 16th International Conference on World Wide Web, Banff, AB, Canada, May 8–12; pp. 271–80. [CrossRef]

Deuze, Mark. 2005. What is journalism? Professional identity and ideology of journalists reconsidered. *Journalism* 6: 442–64. [CrossRef]

Deuze, Mark, and Tamara Witschge. 2018. Beyond journalism: Theorizing the transformation of journalism. *Journalism* 19: 165–81. [CrossRef]

Diakopoulos, Nicholas. 2011. A Functional Roadmap for Innovation in Computational Journalism. Nick Diakopoulos [Blog Post]. Available online: http://www.nickdiakopoulos.com/2011/04/22/a-functional-roadmap-for-innovation-in-computational-journalism/ (accessed on 6 February 2021).

Diakopoulos, Nicholas. 2015. Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism* 3: 398–415. [CrossRef]

Diakopoulos, Nicholas. 2019. *Automating the News: How Algorithms Are Rewriting the Media*. Cambridge: Harvard University Press.

Diakopoulos, Nicholas, and Michael Koliska. 2017. Algorithmic transparency in the news media. *Digital Journalism* 5: 809–28. [CrossRef]

Dörr, Konstantin. Nicholas. 2015. Mapping the field of algorithmic journalism. *Digital Journalism*. [CrossRef]

Fernandez, M., and H. Alani. 2018. Online misinformation: Challenges and future directions. Paper presented at the Companion Web Conference 2018, Lyon, France, April 23–27; pp. 595–602. [CrossRef]

Ford, Martin. 2015. *Rise of the Robots: Technology and the Threat of a Jobless Future*. New York: Basic Books.

Foster, Robin. 2012. *News Plurality in a Digital World*. Oxford: New Reuters Institute for the Study of Journalism.

Frey, Carl Benedikt, and Michael A. Osborne. 2017. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change* 114: 254–80. [CrossRef]

Garrett, R. Kelly. 2009. Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication* 14: 265–85. [CrossRef]

Garrison, Bruce. 1998. *Computer-Assisted Reporting*, 2nd ed. Mahwah: Lawrence Erlbaum Associates.

Gaskins, Benjamin, and Jennifer Jerit. 2012. Internet news: Is it a replacement for traditional media outlets? *The International Journal of Press/Politics* 17: 190–213. [CrossRef]

Glahn, Harry R. 1970. Computer-produced worded forecasts. *Bulletin of the American Meteorological Society* 51: 1126–32. [CrossRef]

Graefe, Andreas. 2016. Guide to automated journalism. *Tow Center for Digital Journalism.* [CrossRef]

Graefe, Andreas, and Nina Bohlken. 2020. Automated Journalism: A Meta-Analysis of Readers' Perceptions of Human-Written in Comparison to Automated News. *Media and Communication* 8: 50–59. [CrossRef]

Grant, Maria J., and Andrew Booth. 2009. A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal* 26: 91–108. [CrossRef]

Haim, Mario, Andreas Graefe, and Hans-Bernd Brosius. 2018. Burst of the filter bubble? Effects of personalization on the diversity of Google News. *Digital Journalism* 6: 330–43. [CrossRef]

Hamilton, James T., and Fred Turner. 2009. Accountability through Algorithm. Center for Advanced Study in the Behavioral Sciences Summer Workshop. Available online: http://web.stanford.edu/~{}fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf (accessed on 5 February 2021).

Hammond, Philip. 2017. From computer-assisted to data-driven: Journalism and Big Data. *Journalism* 18: 408–24. [CrossRef]

Hansen, Mark, Meritxell Roca-Sales, Jon M. Keegan, and George King. 2017. Artificial intelligence: Practice and implications for journalism. In *Tow Center for Digital Journalism*. New York: Columbia University. [CrossRef]

Hong, Hyehyun, and Hyun Jee Oh. 2020. Utilizing bots for sustainable news business: Understanding users' perspectives of news bots in the age of social media. *Sustainability* 12: 6515. [CrossRef]

Jamil, Sadia. 2020. Artificial intelligence and journalistic practice: The crossroads of obstacles and opportunities for the Pakistani journalists. *Journalism Practice*, 1–23. [CrossRef]

Jokela, Sami, Marko Turpeinen, Teppo Kurki, Eerika Savia, and Reijo Sulonen. 2001. The role of structured content in a personalized news service. Paper presented at the 34th Annual Hawaii International Conference on System Sciences, Maui, HI, USA, January 6; pp. 1–10. [CrossRef]

Jones, Bronwyn, and Rhianne Jones. 2019. Public service chatbots: Automating conversation with BBC News. *Digital Journalism* 7: 1032–53. [CrossRef]

Jung, Jaemin, Haeyeop Song, Youngju Kim, Hyunsuk Im, and Sewook Oh. 2017. Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists. *Computers in Human Behavior* 71: 291–98. [CrossRef] [PubMed]

Kennedy, Helen, and Giles Moss. 2015. Known or knowing publics? Social media data mining and the question of public agency. *Big Data & Society* 2. [CrossRef]

Kirley, Elizabeth A. 2016. The robot as cub reporter: Law's emerging role in cognitive journalism. *European journal of Law and Technology* 7: 17–18. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2952151 (accessed on 11 February 2021).

Kitchin, Rob. 2014. Big Data, new epistemologies and paradigm shifts. *Big Data & Society* 1: 1–12. [CrossRef]

Latar, Noam Lemelshtrich. 2015. The robot journalist in the age of social physics: The end of human journalism? In *The New World of Transitioned Media*. Cham: Springer, pp. 65–80. [CrossRef]

Leppänen, Leo, Myriam Munezero, Mark Granroth-Wilding, and Hannu Toivonen. 2017. Data-driven news generation for automated journalism. Paper presented at the 10th International Conference on Natural Language Generation, Santiago de Compostela, Spain, September 4–7; pp. 188–97. [CrossRef]

Lewis, Seth C., Amy Kristin Sanders, and Casey Carmody. 2019. Libel by algorithm? Automated journalism and the threat of legal liability. *Journalism & Mass Communication Quarterly* 96: 60–81. [CrossRef]

Li, Lei, Dingding Wang, Tao Li, Daniel Knox, and Balaji Padmanabhan. 2011. Scene: A scalable two-stage personalized news recommendation system. Paper presented at the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, Beijing, China, July 25–29; pp. 125–34. [CrossRef]

Lindén, Carl-Gustav. 2017. Algorithms for journalism: The future of news work. *The Journal of Media Innovations*. [CrossRef]

Lindén, Carl-Gustav, Hanna Tuulonen, Asta Bäck, Nicholas Diakopoulos, Mark Granroth-Wilding, Lauri Haapanen, Leo Leppänen, Magnus Melin, Tom Moring, Myriam Munezero, and et al. 2019. *News Automation: The Rewards, Risks and Realities of Machine Journalism*. WAN-IFRA Report. Available online: https://cris.vtt.fi/ws/portalfiles/portal/23705408/WAN_IFRA_News_Automation_FINAL.pdf (accessed on 10 February 2021).

Liu, Xiaomo, Armineh Nourbakhsh, Quanzhi Li, Sameena Shah, Robert Martin, and John Duprey. 2017. Reuters tracer: Toward automated news production using large scale social media data. Paper presented at the 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, December 11–14; pp. 1483–93. [CrossRef]

Lokot, Tetyana, and Nicholas Diakopoulos. 2016. News Bots: Automating news and information dissemination on Twitter. *Digital Journalism* 4: 682–99. [CrossRef]

Milosavljević, Marko, and Igor Vobič. 2019. 'Our task is to demystify fears': Analysing newsroom management of automation in journalism. *Journalism*. [CrossRef]

Montal, Tal, and Zvi Reich. 2017. I, robot. You, journalist. Who is the author? Authorship, bylines and full disclosure in automated journalism. *Digital Journalism* 5: 829–49. [CrossRef]

Nechushtai, Efrat, and Seth C. Lewis. 2019. What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior* 90: 298–307. [CrossRef]

Orellana-Rodriguez, Claudia, and Mark T. Keane. 2018. Attention to news and its dissemination on Twitter: A survey. *Computer Science Review* 29: 74–94. [CrossRef]

Otter, Alastair. 2017. Journalism Bots: A Quick History and Ideas for Use in Your Newsroom. Web Article. Available online: https://gijn.org/2017/12/11/journalism-bots-a-quick-history-and-ideas-for-use-in-your-newsroom/ (accessed on 12 February 2020).

Pavlik, John. 2000. The impact of technology on journalism. *Journalism Studies* 1: 229–37. [CrossRef]

Powers, Elia. 2017. My news feed is filtered? Awareness of news personalization among college students. *Digital Journalism* 5: 1315–35. [CrossRef]

Reiter, Ehud. 2010. Natural Language Generation. In *The Handbook of Computational Linguistics and Natural Language Processing*. Edited by Alexander Clark, Chris Fox and Shalom Lappin. Oxford: Wiley-Blackwell, pp. 574–98.

Schapals, Aljosha Karim, and Colin Porlezza. 2020. Assistance or resistance? Evaluating the intersection of automated journalism and journalistic role conceptions. *Media and Communication* 8: 16–26. [CrossRef]

Shao, Chengcheng, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2017. The spread of fake news by social bots. *arXiv* arXiv:1707.07592.

Shin, Jieun, and Thomas Valente. 2020. Algorithms and health misinformation: A case study of vaccine books on amazon. *Journal of Health Communication* 25: 394–401. [CrossRef]

Spyridou, Lia-Paschalia, Maria Matsiola, Andreas Veglis, George Kalliris, and Charalambos Dimoulas. 2013. Journalism in a state of flux: Journalists as agents of technology innovation and emerging news practices. *International Communication Gazette* 75: 76–98. [CrossRef]

Thurman, Neil, Konstantin Dörr, and Jessica Kunert. 2017. When reporters get hands-on with robo-writing: Professionals consider automated journalism's capabilities and consequences. *Digital Journalism* 5: 1240–59. [CrossRef]

Van Dalen, Arjen. 2012. The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism Practice* 6: 648–58. [CrossRef]

Veglis, Andreas, and Efthimis Kotenidis. 2020. Employing chatbots for data collection in participatory journalism and crisis situations. *Journal of Applied Journalism & Media Studies*. [CrossRef]

Veglis, Andreas, and Theodora A. Maniou. 2018. The mediated data model of communication flow: Big data and data journalism. *KOME: An International Journal of Pure Communication Inquiry* 6: 32–43. [CrossRef]

Veglis, Andreas, and Theodora A. Maniou. 2019. Chatbots on the rise: A new narrative in journalism. *Studies in Media and Communication* 7: 1–6. [CrossRef]

Zangana, Abdulsamad. 2017. The Impact of New Technology on the News Production Process in the Newsroom. Ph.D. thesis, University of Liverpool, Liverpool, UK. Available online: https://livrepository.liverpool.ac.uk/3008664/1/201007672_July2017.pdf (accessed on 14 February 2021).

Zhu, Yangyong, Ning Zhong, and Yun Xiong. 2009. Data explosion, data nature and dataology. In *International Conference on Brain Informatics*. Berlin/Heidelberg: Springer, pp. 147–58. [CrossRef]