

What is (automated) news? A content analysis of algorithm-written news articles

Tandoc, Edson C.; Wu, Shangyuan; Tan, Jessica; Contreras-Yap, Sofia

2022

Tandoc, E. C., Wu, S., Tan, J. & Contreras-Yap, S. (2022). What is (automated) news? A content analysis of algorithm-written news articles. *Media E Jornalismo*, 22(41), 103-120. https://dx.doi.org/10.14195/2183-5462_41_6

<https://hdl.handle.net/10356/173664>

https://doi.org/10.14195/2183-5462_41_6

© 2022 Media & Jornalismo. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Downloaded on 11 Sep 2024 22:06:24 SGT

What is (automated) news? A content analysis of algorithm-written news articles

O que são notícias (automatizadas)?
Uma análise de conteúdo de artigos noticiosos redigidos
por algoritmos

https://doi.org/10.14195/2183-5462_41_6

Edson C. Tandoc Jr.

Nanyang Technological University, Singapore
edson@ntu.edu.sg

Shangyuan Wu

National University of Singapore, Singapore
shanwu@nus.edu.sg

Jessica Tan

Nanyang Technological University, Singapore
jessicatan@ntu.edu.sg

Sofia Contreras-Yap

Nanyang Technological University, Singapore
sofiatan001@e.ntu.edu.sg

Submitted: 2022/04/30 | Accepted: 2022/10/10

Submetido: 2022/04/30 | Aceite: 2022/10/10

Abstract

The use of automation in producing news articles confronts journalism with threats, opportunities, and ambiguities. Thus, automation in journalism has attracted a lot of attention, from scholars who sought the perspective of human journalists to those who examined how audiences process algorithm-written news articles. These studies assume that human-written news articles differ from automated-written news articles. But do they? This current study compared human-written with algorithm-written news articles published by media and software company Bloomberg. Guided by the frameworks of field theory and journalistic boundaries, we compared the news articles based on traditional markers of human-written news. Using manual content analysis, we found that algorithm-written news shares some similarities with human-written news, such as focusing on timely or recent events and using the inverted pyramid format. Beyond these, we also found differences. First, in terms of news values, human-written news articles tend to display more negativity and impact than algorithm-written news articles. Human-written news articles are also more li-

kely to include interpretation while algorithm-written articles tend to be shorter and contain no human sources.

Keywords

algorithm, automation, Bloomberg, content analysis, news

Resumo

O uso da automatização na produção de artigos noticiosos confronta o jornalismo com ameaças, oportunidades e ambiguidades. A automatização no jornalismo tem atraído muita atenção por parte da academia, desde a perspectiva dos jornalistas (humanos) à forma como as audiências processam os artigos noticiosos escritos com algoritmos. Estas pesquisas assumem que os artigos noticiosos escritos por humanos diferem dos artigos noticiosos escritos através de processos de automação. Mas será que são diferentes? Este estudo compara os artigos noticiosos escritos por humanos com os artigos noticiosos escritos por algoritmos publicados pela Bloomberg. Mobilizando os enquadramentos da teoria dos campos e a discussão em torno das fronteiras do jornalismo, descobrimos que as notícias escritas através de algoritmos partilham algumas semelhanças com notícias escritas por humanos, como o foco em acontecimentos atuais ou a utilização da pirâmide invertida. Mas também encontrámos diferenças. Primeiro, em termos de valores-notícia, os artigos noticiosos escritos por humanos tendem a exibir mais negatividade e impacto do que os artigos noticiosos escritos por algoritmos. Os artigos noticiosos escritos por humanos são mais suscetíveis de incluir interpretação, enquanto os artigos escritos por algoritmos tendem a ser mais curtos, sem utilizar fontes humanas.

Palavras-chave

algoritmos, automatização, Bloomberg, análise de conteúdo, notícias

Introduction

Technology has always been a transformative force in journalism (Pavlik, 2000). Major innovations, like the production of mobile, immersive, and data stories brought about by new technologies, have restructured work processes and reshaped journalistic outputs (Lewis & Zamith, 2017; Pavlik, 2000; Tandoc, 2019). While technological innovations may appear to be smoothly integrated in the news production process, tensions manifest due to various concerns regarding technology's impact on ethics, job security, and information integrity (Flew et al., 2012; Lewis & Westlund, 2015; Van Dalen, 2012; Weaver & Willnat, 2016). One innovation in journalism where such concerns are specifically salient is automated journalism.

The use of automated journalism, which some narrowly define as referring to computer-written news with little to no human input beyond initial programming (Carlson, 2015; Linden, 2017), confronts the journalistic field with threats, opportunities, and ambiguities. For some, automated newswriting frees up human journalists from writ-

ing trivial, templated articles and allows them to focus on writing news that requires more analysis and higher discernment. For others, introducing automated newswriting in the newsroom presents yet another challenge to human journalists' ethical stance, editorial control, and job security, which are already all under threat. It presents ambiguities, as while organizations may benefit from a more efficient news production process, individual journalists fear job displacement or the need for retraining. Thus, automation in journalism has attracted a lot of scholarly attention, from those who sought the perspective of human journalists on how automation is unfolding in the field, to those who examined how audiences process algorithm-written news articles. To a large extent, many of these studies assume that human-written news articles differ from automated-written news articles. But do they?

This current study seeks to answer this question by comparing human-written and algorithm-written news articles published by financial media company Bloomberg, one of the early adopters of automated newswriting. Bloomberg publishes algorithm-written news articles under the byline of "Bloomberg Automation." Through the lens of Bourdieu's (1998, 2005) field theory and a content analysis of 1,282 randomly selected news articles published by Bloomberg from 2016 to 2017, this study compares human-written and algorithm-written articles based on established benchmarks of traditional news, such as the presence of particular news values, the use of sources and typical formats, and the presence of interpretation, and discusses the possible impact of automated newswriting's growth on the field of journalism.

Automated journalism

Seen as a subset of "computational journalism" or "algorithmic journalism" (Lewis and Zamith, 2017; Lewis et al. 2019), automated journalism has been initially defined as computer-written news with little to no human input beyond initial programming (Carlson, 2015; Linden 2017), although others later argued that automation in journalism goes beyond just writing, but also data collection and management, among other news production processes (Wu et al., 2018). Thus, automated journalism involves the use of automation technologies, which perform various tasks like writing, information filtering, and classification (Diakopoulos, 2015), in any part of the news production process, from the news gathering and selection stages to the news writing, editing, and distribution stages (Wu et al., 2018). Technological innovations, such as artificial intelligence software programs capable of machine learning and natural language generation, facilitate automated journalism (Montal and Reich, 2016). This reduction in human labor translates to reduced costs in news production (Carlson, 2015; Fanta, 2017), which may have contributed to the widespread adoption of automated journalism among news organizations like *The Washington Post*, the Associated Press, and the BBC (Danzon-Chambaud & Cornia, 2021), especially for data-driven news articles, such as those about earthquakes, sports, and business.

With automation's increasing ubiquity in the news production process, researchers have been studying its influence on the practices, content, business models, and labor requirements of traditional news media organizations. Carlson (2015) found both positive (e.g., finding patterns in information usually missed by humans) and

negative (e.g., increased layoffs of journalists) outcomes with automation technology use in newsrooms. Two studies raised ethical concerns regarding the use of algorithms and automation technology: Parasie (2015) found that journalists encountered dilemmas on their duty toward readers regarding information accuracy, and Montal and Reich (2016) discovered a lack of disclosure and byline policies for automated journalism articles. Van Dalen (2012) also found that automation lowers the costs and increases the efficiency in newsrooms and enables human journalists to pursue more challenging and creative stories.

Some news outlets have adopted automation in their processes. For example, Linden (2017) found that weather, sports, medicine, and business and finance news articles lend themselves better to automation. In particular, news professionals have described business news as a genre that is “easily templated” and involves frequent “repetition” (Wu et al., 2018, p. 11). For example, in commodities market news, automation technology allows the easy checking and reporting of commodity price movements on a daily basis with zero involvement of a human journalist (Wu, 2020). High speed information scanning and story production allow automated journalism to match the pace of trading decisions, which are also left to “software designed to find marginal advantages and a competitive edge at speeds no human can replicate” (CB Insights, 2016, p. 11). This may be one reason behind financial news wires’ well-publicized adoption of automated journalism. For example, financial, software, and media company Bloomberg has established a separate “fully automated news service” called Bloomberg Automated Intelligence (BAI) (Fesanghary & Verma, 2020, p. 2).

Bloomberg, a top provider of business news with over 2,700 journalists and analysts producing over 5,000 articles every day (Bloomberg, n.d.), has widely used automation to produce news articles. Its BAI service leverages Bloomberg’s extensive datasets, tracks markets, finds useful data, and shares all this information through automated computer-written articles created from more than 500 story templates (Bloomberg, 2021). In a speech at the Digital Life Design Conference in early 2019, Bloomberg Editor-in-Chief John Micklethwait revealed that Bloomberg’s Cyborg bot can swiftly extract key details from earnings reports to generate headlines and short news articles. He also disclosed that about a third of all Bloomberg news content is produced with some form of automation (Digital-Life-Design, 2019). Bloomberg’s status as a leading provider of global financial news and its extensive use of automation in producing news articles provides a useful and important context within which the impact of automation on journalism may be investigated. Thus, in this current study, we examine Bloomberg’s use of automation in news production by comparing its algorithm-written news articles with those written by its human journalists. This current study focuses on automated newswriting, which is only one aspect of automated journalism.

Transformations in the journalistic field

Pavlik (2000, p. 229) argued that journalism “has always been shaped by technology.” Indeed, most recent developments in the journalistic field, such as mobile, immersive, and automated journalism, are the results of technological innovation (Lewis

& Zamith, 2017). While many of these technologies were not specifically designed for journalism, their use in journalism can be transformative, as they may bring logics and practices not originally intended for journalistic purposes (Tandoc, 2019). Research interest in the increasing use of different technologies in journalism has naturally followed, with studies investigating technology's effect on news production. For example, Perreault and Stanfield (2019) examined role conceptions in mobile journalism; Danzon-Chambaud and Cornia (2021) examined automated journalism's impact on media practitioners; and Fahmy and Attia (2020) examined data journalism practice and development in the Arab world. What is common among these studies examining technology's impact on journalism is their use of Bourdieu's (1998, 2005) field theory to guide and frame their research. In field theory, journalism is seen as a field of forces that may be transformed or preserved depending on the actions and decisions of existing agents and new entrants (Bourdieu, 2005). New players entering the field also introduce new processes (e.g., web analytics in journalism), which have been studied to reveal how a technology developed outside the field of journalism is changing traditional journalistic norms and routines (e.g., Wang, 2018; Moyo et al., 2019; Hanusch, 2017; Petre, 2018; Wu et al., 2019a).

The field of journalism can be susceptible to changes brought about by external shocks due to its highly heterogeneous capital structure (Bourdieu, 1998; 2005). Not only does it capitalize on its cultural capital that is often operationalized as the field's cache of competence and credibility, but the journalistic field also relies heavily on its economic capital, usually measured in terms of revenues and audience size (Benson & Neveu, 2005). The field's embrace of new technologies, including web analytics and automation, has been widely seen as a response to journalism's shrinking stock of economic capital (Tandoc, 2014; Dörr, 2016). When new technologies, processes, and agents enter the journalistic field, they bring with them logics that are external to the field but are then able to challenge, if not transform, journalism's internal logics (Wu et al., 2019b; Belair-Gagnon & Holton, 2018; Eldridge, 2018).

Indeed, automation is a process that originated from outside the field of journalism (Danzon-Chambaud & Cornia, 2021; Linden, 2017). How automation impacts the content of news has been discussed mostly from the perspective of audiences and pertaining to perceptions of credibility and objectivity. Waddell (2018) discovered that audiences view articles declared to be written by human journalists to be more credible than those declared to be written by algorithms; Liu and Wei (2019) found audiences perceiving machine written news to be more objective than human-written news; and Waddell (2019) found that articles co-authored by human and machine are viewed as less biased by audiences than those written just by the machine. Tandoc et al. (2020) also discovered that when audiences perceive a story's content to be objective, they tend to rate message and source credibility higher if it was written by a machine than by a human. However, there has not been any empirical research done on the actual differences in content produced by a machine versus that produced by a human journalist.

Studies have compared online news articles and print news articles, finding some differences, such as online articles more likely to engage in follow-up reporting than do print articles (Burggraaff & Trilling, 2017). Comparing content written through automation with those written by a human journalist is also an important investigation,

given assertions by scholars that algorithm-written output may be more scientific, precise, and neutral because of its seeming lack of personal bias or opinion (Parasie, 2015; Borges-Rey, 2016; Tandoc & Oh, 2017). Conversely, humans have been perceived to be able to produce content that machines cannot because they are able to conduct further inquiry, use critical thought and observation, and perform in-depth analysis and investigation (Abu-Fadil, 2016). Tasks like conducting interviews; injecting emotion, wit, and insight into articles; recognizing political, legal, and cultural sensitivities; and establishing context and causality still lie within the purview of human journalists and highlight the importance of their contributions (Wu et al., 2018). These manifest in human journalists' ability to inject opinion, analysis, and context into their reporting, which can be considered as engaging in interpretation, going beyond the dissemination of information. However, the extent to which automated news articles may be adhering to traditional journalistic rules that have guided human journalists in their writing has not been adequately studied.

Boundaries of journalism

While field theory may provide a broad argument to why automation in the news must be studied as it brings in logics originally developed outside the field, the concept of journalistic boundaries is also instructive. Increasingly, actors who would not fall under the traditional definition of a journalist—such as ordinary citizens or data scientists—are now performing acts of journalism and have been referred to as interlopers or peripheral actors in journalism (Belair-Gagnon & Holton, 2018; Eldridge, 2018). Traditional forms of news dissemination have been upended by social media and messaging apps (Kim, 2020; Bosch, 2014), and traditional ways of writing are now supplemented by automated news writing (Jung et al., 2017; Liu & Wei, 2019; Montal & Reich, 2017; Wu et al., 2019a; Tandoc et al., 2020). Thus, journalists find themselves having to constantly negotiate the boundaries of their profession (Carlson, 2015).

What are the boundaries of journalism? In proposing a theory of metajournalistic discourse, which refers to “public expressions evaluating news texts, the practices that produce them, or the conditions of their reception,” Carlson (2016, p. 362) identified boundary setting as an important process and called attention to “boundaries around actors, norms, and practice.” The concept of boundary work traces to Gieryn’s (1983) observation of how scientists engage in strategies to demarcate or contrast their work from non-scientific or technical pursuits to maintain a particular public image. Studies in journalism have since adapted the concept to describe how traditional journalistic actors distinguish themselves from external actors or new entrants. For example, in their analysis of metajournalistic discourse around Gawker’s outing of a married magazine executive, Tandoc and Jenkins (2016) found that news outlets and reader comments focused on outlining who a journalist is, what constitutes news, as well as ethical standards as important boundaries of journalism. In further operationalizing journalism’s boundaries to analyze big data journalism, Tandoc and Oh (2017) conducted a content analysis of *The Guardian’s* Datablog articles based on news values, sources, topics, visualization, and objectivity. Stalph (2018) also examined data-driven articles and focused on some of these markers of traditional jour-

nalism, classifying them into formal characteristics (e.g., number of words, topic); data visualization (e.g., visualization type); data sources (e.g., data provider, country of origin); and form and content (e.g., story format, subject matter). In their analysis of data-driven news outputs in China, Zhang and Feng (2019) also focused on data source, data analysis, mode of presentation, and transparency.

Guided by these studies and the frameworks of field theory and boundary work, this current study compares automated news articles with articles written by human journalists from Bloomberg, one of the pioneers in automated news writing, in terms of the articles' adherence to journalism's news values and routines. Studies have long examined and considered these news values and routine manifestations as markers of traditional news, with the assumption that traditional news is produced, or constructed, by human journalists. *News values* have been described as a set of "rules" (Shoemaker & Vos, 2009, p. 53) and a set of "requirements" (Harcup & O'Neill, 2001, p. 1471) that guide what messages and information are emphasized and selected as news (Tandoc et al., 2021). They influence journalists' decision-making throughout the whole news production process (Parks, 2019). Common news values used in studies include timeliness, proximity, impact, and novelty, among others (Cagle & Bednarek, 2015). In their content analysis of articles based on big data published in *The Guardian's* Datablog, Tandoc and Oh (2017) found that most articles displayed the news value of prominence, using datasets from well-known organizations and featuring prominent companies and countries in their reports.

News routines are "repeated practices and forms" that allow journalists to efficiently perform news production within temporal and economic constraints (Lowrey, 2014, p. 1). News routines can also manifest in news outputs, such as in topic selection and use of particular sources. Journalists, dealing with constraints like deadlines, viewership levels, and information availability, may be limited to covering stories under certain topics only, or relying on a small subset of usual news sources. For example, a content analysis of online news articles from 10 news sites in five Western countries found a strong emphasis on news about politics and economy (Quandt, 2008). A content analysis of broadcast, print, and online news from 11 countries had found an over-reliance on government sources in most countries (Tiffen et al., 2014). Another news routine is the use of the inverted pyramid as story format, which Pöttker (2003, p. 510) describes as "cost-saving" as it allows quicker editing and faster production of articles. A comparative content analysis of news articles published by *The New York Times* and BuzzFeed in the United States found that majority of their articles (82% for BuzzFeed, 71% for *The New York Times*) used the inverted pyramid format (Tandoc, 2018).

Based on these previous studies that examined news articles based on traditional content markers, as well as on the ongoing discourse and research about the limitations of automated news, this current study focuses on Bloomberg and compares its news articles written by humans with those written by automated technology based on the following content features: news values, topic, sources, format, and providing context and analysis. Therefore, using manual content analysis, we seek to answer the following research question:

- RQ. How do human-written and algorithm-written news articles compare in terms of:
1. News values (i.e., negativity, timeliness, impact, novelty, superlativeness)?

2. News topic?
3. Dominant information source?
4. Story format?
5. Providing interpretation (i.e., opinion, context, and analysis)?

Method

This study is based on manual content analysis of news articles published by Bloomberg across two years: 2016 and 2017. The unit of analysis is each individual news article. Content analysis allows the examination of the extent to which content elements recur, but this is limited to only manifest content (i.e., what is in the data), which potentially misses out on analyzing nuances and sensemaking that qualitative approaches engage in. Launched in the 1980s, Bloomberg has long been known for its information “terminal,” a computer software system designed to meet the needs of finance professionals such as bankers, analysts and traders who require real-time and newsworthy information related to the economy and financial markets. While it provides news and data services across other platforms like the internet, radio, print, and television, Bloomberg’s terminal offerings remain a core business (Stewart, 2019), with more than 320,000 subscribers worldwide (CB Insights, 2016). Having established itself as a leading provider of global financial news, Bloomberg serves as a noteworthy case study in automated newswriting. The company has increasingly focused on its capacities in automated journalism, a domain wherein competitors such as Associated Press, *The Washington Post* and Reuters are also developing expertise.

Using constructed week sampling (Riffe et al., 1993), our analysis involves 1,280 articles: 650 articles are human-written articles, and 630 articles are algorithm-written articles. We sourced the articles from our university’s subscription to Bloomberg’s computer terminal. This paid subscription provides access to Bloomberg’s news reports, among other types of content. In this study, we focused on news articles in the terminal. We collected a representative sample of articles using constructed week sampling by constructing two full weeks for each year. To construct one full week for 2016, we randomly selected one Monday, one Tuesday, and so forth (e.g., using random selection, we randomly picked one Monday and then sampled all articles published on that day). We repeated this process to construct a second week for 2016, ensuring that no two dates are repeated (Riffe et al., 1993). The same process was done to sample articles from 2017.

Based on the literature, we developed a coding manual and trained three student coders in using the manual to analyze the sampled articles. Following the training, an intercoder testing exercise was conducted, where the coders independently analyzed 20 articles, evenly split between human and automated and excluded from the final sample. While intercoder reliability among the three coders was achieved in most categories, issues were found for some variables, including coding for the dominant source (Krippendorff’s $\alpha = .62$) and the news value of proximity (Krippendorff’s $\alpha = .32$). Proximity is defined as the geographical or psychological closeness of an issue or event to the newsroom (Martin, 1988); however, Bloomberg operates several bureaus which is difficult to ascertain based on content alone. Thus, we decided to

exclude this variable in our analysis. We conducted another round of coder training and subsequently conducted another round of intercoder testing, which yielded acceptable intercoder reliability results (Krippendorff's $\alpha = .70$ and above), allowing us to proceed with the actual coding of the articles based on the following categories:

Author type. First, we coded the articles if they were written by a human, using automation, or a combination of both human and automation. Bloomberg labels automated news articles with a byline that says "Bloomberg Data News" as well as a sentence at the end of the article stating that: "This story was produced by the Bloomberg Automated News Generator." Since this was a straightforward categorization, and no article had combined authorship, the coders achieved perfect reliability (Krippendorff's $\alpha = 1.0$).

News values. The articles were coded for the presence or absence of five news values common in traditional journalism (Caple and Bednarek, 2015; Harcup and O'Neill, 2001; Rogers, 2004). Negativity refers to when the article focuses on the negative aspects of an issue or event (Krippendorff's $\alpha = 0.74$). Timeliness refers to whether the news article is about a recent issue or event, and we coded each article based on its date of publication, so that an article is coded as having the news value of timeliness if the event or issue mentioned in the article is temporally close to the publication date (Krippendorff's $\alpha = 1.0$). Impact refers to whether the article is about an issue or event that has significant effect or consequences on a large number of people (Krippendorff's $\alpha = 0.71$). Novelty refers to whether the issue or event depicted in the article has new or unexpected aspects (Krippendorff's $\alpha = 1.0$). Finally, superlativeness refers to whether the issue or event itself has a large scope or scale, such as involving a high number of participants, etc. (Krippendorff's $\alpha = 0.74$). Eliteness or prominence was excluded from the analysis as the coders yielded consistently low reliability in both rounds of testing.

Topic. We adopted Bloomberg's own categorization of topics as specified in the terminal. Thus, the articles were coded if they were about business news, general news, legal affairs, or sports. The coders recorded a reliability level slightly lower than the study's threshold (Krippendorff's $\alpha = 0.69$)—this limitation must be considered when interpreting the results.

Dominant source. The articles were coded for the dominant human source used. We only focused on the main source, based on who is quoted in the lead or cited in most paragraphs. We also decided to narrow down to just the human source, although an article can also cite document sources. The dominant source can be from government, politics, or law enforcement; business; civic society; culture, arts, sports, or entertainment; academe; or ordinary people. We also coded for the absence of any human source. Following a second training session and intercoder reliability testing, reliability improved from the first round (Krippendorff's $\alpha = .90$).

Story format. This refers to the story structure, which can be in the form of a listicle, chronology, reverse chronology, narrative, or the commonly used inverted pyramid (Tandoc, 2018). The coders achieved perfect reliability. We also coded for story length based on number of words, where the coders achieved acceptable reliability (Krippendorff's $\alpha = .78$).

Providing interpretation. Based on previous studies that argued about the limitations of automated news and the advantages of human authors, we also coded the

articles on whether they include opinion (e.g., personal interpretation of an issue or event), context (e.g., background information or description of the bigger picture), and analysis (e.g., explanations or critical perspectives on the event). An article got a score of 1 for each of these elements; the scores across the three elements were added, so that an article can get a maximum score of 3 if all these elements were present. The coders achieved acceptable reliability (Krippendorff's $\alpha = .74$).

Results

Using manual content analysis, we compared Bloomberg articles written by human authors with those written by its automated news generator based on news values, topic, source, format, and interpretation, which are usually considered markers of traditional news, long been assumed to be produced by humans.

First, we focus on news values: negativity, timeliness, impact, novelty, and superlativeness. The analysis found that all articles analyzed contained the news values of timeliness and novelty, except for one algorithm-written article. This may be due to the nature of Bloomberg's subscription-based news service that focuses on sending out news quickly to subscribers. Thus, we proceed to the three other news values: negativity, impact, and superlativeness. The analysis found significant associations between type of author (human vs. machine) and the presence of the three news values (Table 1).

Table 1 – Comparing human-written and algorithm-written articles

Categories		Human-Written	Algorithm-written	χ^2
News values	Timeliness	100%	99.8%	1.03
	Novelty	100%	99.8%	1.03
	Negativity	49.1%	27.5%	63.18 a
	Impact	20.3%	5.7%	59.93 a
	Superlativeness	21.2%	5.6%	67.25 a
News topic	Business news	47.3%	97.9%	408.37a
	General news	52.7%	2.1%	
Source	Human source	93.5%	0.6%	1108.60a
Format	Inverted pyramid	96.2%	99.8%	22.86a
	Opinion	21.5%	6%	64.25 a
Interpretation	Context	96.5%	100%	22.39 a
	Analysis	36.3%	6.8%	162.21 a

Note. ^a $p < .001$.

In terms of negativity ($\chi^2[1, N = 1280] = 63.18, p < .001$), human-written articles (49.1%) had a higher likelihood of containing negativity than algorithm-written articles (27.5%). Next, in terms of impact ($\chi^2[1, N = 1280] = 59.93, p < .001$), human-writ-

ten articles (20.3%) had a higher likelihood of containing the news value of impact than algorithm-written articles (5.7%). Finally, in terms of superlativeness $t(x^2[1, N = 1280] = 67.25, p < .001)$, human-written articles (21.2%) had a higher likelihood of containing the news value of superlativeness than algorithm-written articles (5.6%).

Second, the analysis found significant association between type of author (human vs. automation) and broad topic, based on Bloomberg's own categorization, $(x^2[1, N = 1280] = 408.37, p < .001)$; see Table 1). Human-written articles were almost evenly split into business news (48.3%) and general news (52.7%) while algorithm-written news was mostly about business news (97.9%).

Third, the analysis found significant association between type of authorship and the use of human source $(x^2[1, N = 1280] = 1108.61, p < .001)$; see Table 1). While we have coded for specific types of sources, we had to dummy code the variable into with human vs no human source to be able to run meaningful chi-square test of association, as algorithm-written articles were mostly devoid of human sources. While human-written articles were predominantly relying on human sources (93.5%) such as business-people, almost all algorithm-written articles did *not* mention a human source (99.4%).

Fourth, in terms of story format, most human-written articles used the inverted pyramid format (96.2%) while all the algorithm-written articles except one (99.8%) used inverted pyramid (see Table 1). In terms of story length, there was a significant difference between human-written and algorithm-written articles, $t(1278) = 23.61, p < .001$. Human-written articles ($M = 476, SD = 296$) were significantly longer than algorithm-written articles ($M = 189, SD = 72$).

Finally, when it comes to providing interpretation—inclusion of opinion, context, and analysis—we found a significant, albeit small, difference between human-written and algorithm-written articles, $t(1258) = 11.11, p < .001$. Human-written articles were slightly more likely to exhibit various types of interpretation ($M = 1.54, SD = .82$) than algorithm-written articles ($M = 1.13, SD = .44$). The smaller variation in scores among algorithm-written articles also shows that algorithm-written articles tend to be more uniform, while human-written articles exhibited more variability in the extent to which forms of interpretation are included (see Table 1 for a detailed comparison).

Discussion

This current study compared human-written with algorithm-written news articles published by Bloomberg and archived in its terminal, which supplies a range of content, including news articles, to subscribers. Guided by the frameworks of field theory and journalistic boundaries, we compared the news articles based on traditional markers of news, which have been examined for a long time with the assumption that news is produced and constructed by human journalists. By using these traditional markers, we found that algorithm-written news shares some similarities with human-written news, such as focusing on timely or recent events and using the inverted pyramid format. Beyond these, we also found differences. First, human-written news articles tend to display more negativity and impact than algorithm-written news articles. Human-written news articles are also more likely to include interpretation, which many scholars studying automation in news have

argued based on interviews with human journalists. Algorithm-written articles also tend to be shorter in length and contain no human sources.

Automation in the news has been initially welcomed for its promise of efficiency and speed. Indeed, we see this in the news value of timeliness, that algorithm-written news can assemble details faster, using automated processes. Since they are automatically generated using algorithms and templates that were initially programmed by humans, they also usually contain news elements that are easily templated, such as story format. These findings show that machines can be programmed to mimic human output, at least to some extent. Across other aspects, however, we still find divergence between human-written and algorithm-written outputs.

First, we found that human-written news articles are more likely to provide interpretation. But a closer scrutiny of the data also reveals that algorithm-written articles also provide background information, a form of context (see Table 1). Background information can also be templated, at least in business news, such as providing annual financial trends to contextualize daily reports. But human-written articles contain more analysis and opinion. Indeed, unlike machines, human journalists can inject opinion, analysis, and context into their work (Wu et al., 2018). Second, and closely related to the earlier point, human-written articles tend to be longer, which can be explained by the injection of analysis and opinion. Differences in length may also be due to the use of human sources, common in human-written articles but absent in algorithm-written articles. Third, we found that Bloomberg uses automation almost exclusively for business news, while human-written articles represent a wider range of topics. This may be representing what previous studies have argued, that automation in journalism can free up human journalists to pursue more important, and perhaps more diverse, stories.

These findings also seem to show automation, at least in Bloomberg, is kept in its place by human managers—the humans continue to control the types of articles the machines are tasked to write, overseeing their work, and perhaps ensuring that they do not take on tasks that significantly threaten the position of human journalists in the newsroom hierarchy. Machines are delegated articles that are number-oriented and easily templated, i.e., business news rather than general news, and that have less impact and are of a smaller scope or scale. These algorithm-written articles tend also to be devoid of human perspectives and have less interpretation. Human managers in the newsroom, it seems, are still controlling the rate at which machines are *allowed* to transform the news industry. That said, if automated news writing continues to grow in scale and improve in its ability to mimic human writing, how will it affect variety and diversity in news coverage? For example, our findings show that while human-written articles have a lot of variance when it comes to length and injection of interpretation; algorithm-written articles tend to be more uniform, displaying less variety. News organizations keen to utilize such automated technologies to reduce their dependence on manpower and increase their output may actively make decisions to machine-write a greater number of articles. Concerns then may arise, as news produced by machines tend to contain less critical thought, inquiry, and investigation (Abu-Fadil, 2016). Output quantity may come at the expense of quality.

Of course, the findings presented here must be understood in the context of several limitations. First, while manual content analysis allows empirical comparison

of textual elements, it cannot capture the contexts and processes behind the production of manifest content. For example, our results cannot ascertain what kinds of editorial policies are in place that guide automated news writing (e.g., what kinds of stories are “assigned” to algorithms). Still, the findings presented here can complement interview-based studies that explored what journalists think about and do with automation. Second, we focused on a specific news organization, one that has the resources to experiment with and fine-tune the embedding of automation in its work processes, which limits the applicability of the insights gleaned from our analysis. Future studies can build on our findings to examine the content produced by automated processes implemented in other types of news organizations as well as articles. For example, other news organizations use automated newswriting for earthquake articles and sports news articles. Finally, our sample of news articles was from 2016 to 2017, and the use of automated newswriting may have changed since then, both within and outside Bloomberg, especially during the COVID-19 pandemic, where a lot of news coverage focused on pandemic-related statistics. Still, we hope the findings we presented here will be useful for future studies as we continue to examine and understand how automation is (or is not) transforming the journalistic field.

References

- Abu-Fadil, M. (2016, September 25). *Will automation upend journalism?* Huffington Post. https://www.huffpost.com/entry/will-automation-upend-jou_b_12179988
- Belair-Gagnon, V., & Holton, A.E. (2018). Boundary work, interloper media, and analytics in newsrooms. *Digital Journalism*, 6(4), 492-508. <https://doi.org/10.1080/21670811.2018.1445001>
- Benson, R., & Neveu, E. (2005). Introduction: Field theory as a work in progress. In R. Benson & E. Neveu (Eds.), *Bourdieu and the journalistic field* (pp.1–28). Polity Press.
- Bloomberg. (n.d.). *What we do. Careers*. Bloomberg. <https://www.bloomberg.com/company/what-we-do/>
- Bloomberg. (2021). *Using Bloomberg automated news stories to predict market events*. Bloomberg. <https://www.bloomberg.com/professional/blog/using-bloomberg-automated-news-stories-to-predict-market-events>
- Borges-Rey, E. (2016). Unravelling data journalism: A study of data journalism practice in British newsrooms. *Journalism Practice*, 10(7), 833–843. <https://doi.org/10.1080/17512786.2016.1159921>
- Bosch, T. (2014). Social media and community radio journalism in South Africa. *Digital Journalism*, 2(1), 29–43. <https://doi.org/10.1080/21670811.2013.850199>
- Bourdieu, P. (1998). *On television*. The New Press.
- Bourdieu, P. (2005). The political field, the social science field and the journalistic field. In R. Benson & E. Neveu (Eds.), *Bourdieu and the journalistic field* (pp. 29–47). Polity Press.
- Burggraaff, C., & Trilling, D. (2017). Through a different gate: An automated content analysis of how online news and print news differ. *Journalism*, 21(1), 112–129. <https://doi.org/10.1177/1464884917716699>
- Caple, H., & Bednarek, M. (2015). Rethinking news values: What a discursive approach can tell us about the construction of news discourse and news photography. *Journalism*, 17(4), 435–455. <https://doi.org/10.1177/1464884914568078>

- Carlson, M. (2015). Introduction: The many boundaries of journalism. In M. Carlson & S.C. Lewis (Eds.), *Boundaries of journalism: Professionalism, practices, and participation* (pp. 1-18). Routledge.
- Carlson, M. (2016). Metajournalistic discourse and the meanings of journalism: Definitional control, boundary work, and legitimation. *Communication Theory*, 26(4), 349–368. <https://doi.org/10.1111/comt.12088>
- CB Insights. (2016). *Twilight of the terminal: Disruption of Bloomberg LP*. <https://www.cbinsights.com/research/report/bloomberg-terminal-disruption/>
- Danzon-Chambaud, S., & Cornia, A. (2021). Changing or reinforcing the “rules of the game”: A field theory perspective on the impacts of automated journalism on media practitioners. *Journalism Practice*, 22(14), 1987-2004. <https://doi.org/10.1080/17512786.2021.1919179>
- Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital journalism*, 3(3), 398-415. <https://doi.org/10.1080/21670811.2014.976411>
- Digital-Life-Design. (2019, January 21). *Journalism in the age of AI (John Micklethwait, Bloomberg Media) | DLD 19* [Video]. YouTube. <https://www.youtube.com/watch?v=65jDYCAnLJU>
- Dörr, K.N. (2016). Mapping the field of algorithmic journalism. *Digital Journalism*, 4(6), 700-722. <https://doi.org/10.1080/21670811.2015.1096748>
- Eldridge, S.A. (2018). “Thank God for Deadspin”: Interlopers, metajournalistic commentary, and fake news through the lens of “journalistic realization”. *New Media & Society*, 21(4), 856-878. <https://doi.org/10.1177/1461444818809461>
- Fahmy, N., & Attia, M. A. M. (2021). A field study of Arab data journalism practices in the digital era. *Journalism Practice*, 15(2), 170-191. <https://doi.org/10.1080/17512786.2019.1709532>
- Fanta, A. (2017). *Putting Europe’s robots on the map: Automated journalism in news agencies*. Reuters Institute for the Study of Journalism. <https://reutersinstitute.politics.ox.ac.uk/our-research/putting-europes-robots-map-automated-journalism-news-agencies>
- Fesanghary, M., & Verma, A. (2021). *Predictive analysis of Bloomberg Automated Intelligence*. Bloomberg Quant Research. <https://www.bloomberg.com/professional/blog/using-bloomberg-automated-news-stories-to-predict-market-events/>
- Flew, T., Spurgeon, C., Daniel, A., & Swift, A. (2012). The promise of computational journalism. *Journalism practice*, 6(2), 157–171. <https://doi.org/10.1080/17512786.2011.616655>
- Gieryn, T. F. (1983). Boundary-Work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American Sociological Review*, 48(6), 781–795. <https://doi.org/10.2307/2095325>
- Hanusch, F. (2017). Web analytics and the functional differentiation of journalism cultures: individual, organizational and platform-specific influences on newswork. *Information, Communication & Society*, 20(10), 1571–1586. <https://doi.org/10.1080/1369118X.2016.1241294>
- Harcup, T., & O’Neill, D. (2001). What is news? Galtung and Ruge revisited. *Journalism studies*, 2(2), 261–280. <https://doi.org/10.1080/14616700118449>
- Jung, J., Song, H., Kim, Y., Im, H., & Oh, S. (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, 291–298. <https://doi.org/10.1016/j.chb.2017.02.022>
- Kim, H.S. (2020). How message features and social endorsements affect the longevity of news sharing. *Digital Journalism*, 9(8), 1162-1183. <https://doi.org/10.1080/21670811.2020.1811742>
- Lewis, S. C., Guzman, A. L., & Schmidt, T. R. (2019). Automation, journalism, and human-machine communication: Rethinking roles and relationships of humans and machines in news. *Digital Journalism*, 7(4), 409–427. <https://doi.org/10.1080/21670811.2019.1577147>

- Lewis, S.C., & Westlund, O. (2015). Big data and journalism: Epistemology, expertise, economics, and ethics. *Digital journalism*, 3(3), 447–466.
<https://doi.org/10.1080/21670811.2014.976418>
- Lewis, S.C., & Zamith, R. (2017). On the worlds of journalism. In P.J. Boczkowski & C.W. Anderson (Eds.), *Remaking the news: Essays on the future of journalism scholarship in the digital age* (pp.111–128). The MIT Press. <https://doi.org/10.7551/mitpress/10648.003.0012>
- Linden, C. G. (2017). Decades of automation in the newsroom: Why are there still so many jobs in journalism? *Digital journalism*, 5(2), 123–140.
<https://doi.org/10.1080/21670811.2016.1160791>
- Liu, B., & Wei, L. (2019). Machine authorship in situ: Effect of news organization and news genre on news credibility. *Digital Journalism*, 7(5), 635–657.
<https://doi.org/10.1080/21670811.2018.1510740>
- Lowrey, W. (2014). News routines. In W. Donsbach (Ed.), *The international encyclopedia of communication*. <https://doi.org/10.1002/9781405186407.wbiecn028>
- Martin, S.R. (1988). Proximity of event as factor in selection of news sources. *Journalism Quarterly*, 65(4), 986–989. <https://doi.org/10.1177/107769908806500424>
- Montal, T., & Reich, Z. (2017). I, robot. You, journalist. Who is the author? Authorship, bylines and full disclosure in automated journalism. *Digital Journalism*, 5(7), 829–849.
<https://doi.org/10.1080/21670811.2016.1209083>
- Moyo, D., Mare, A. & Matsilele, T. (2019). Analytics-driven journalism? Editorial metrics and the reconfiguration of online news production practices in African newsrooms. *Digital Journalism*, 7(4), 490–506. <https://doi.org/10.1080/21670811.2018.1533788>
- Parasie, S. (2015). Data-driven revelation? Epistemological tensions in investigative journalism in the age of “big data”. *Digital Journalism*, 3(3), 364–380.
<https://doi.org/10.1080/21670811.2014.976408>
- Parks, P. (2019). Textbook news values: Stable concepts, changing choices. *Journalism & Mass Communication Quarterly*, 96(3), 784–810. <https://doi.org/10.1177/1077699018805212>
- Pavlik, J. V. (2000). The impact of technology on journalism. *Journalism Studies*, 1(2), 229–237. <https://doi.org/10.1080/14616700050028226>
- Perreault, G., & Stanfield, K. (2019). Mobile journalism as lifestyle journalism? Field theory in the integration of mobile in the newsroom and mobile journalist role conception. *Journalism Practice*, 13(3), 331–348. <https://doi.org/10.1080/17512786.2018.1424021>
- Petre, C. (2018). Engineering consent: How the design and marketing of newsroom analytics tools rationalize journalists’ labor. *Digital Journalism*, 6(4), 509–527.
<https://doi.org/10.1080/21670811.2018.1444998>
- Pöttker, H. (2003). News and its communicative quality: the inverted pyramid—when and why did it appear? *Journalism Studies*, 4(4), 501–511. <https://doi.org/10.1080/1461670032000136596>
- Quandt, T. (2008). (No) news on the world wide web? *Journalism Studies*, 9(5), 717–738.
<https://doi.org/10.1080/14616700802207664>
- Riffe, D., Aust, C. F., & Lacy, S. R. (1993). The effectiveness of random, consecutive day and constructed week sampling in newspaper content analysis. *Journalism Quarterly*, 70(1), 133–139.
<https://doi.org/10.1177/107769909307000115>
- Rogers, T. (2004). *Newswriting on deadline*. Pearson.
- Stalph, F. (2018). Classifying data journalism: A content analysis of daily data-driven stories. *Journalism Practice*, 12(10), 1332–1350. <https://doi.org/10.1080/17512786.2017.1386583>
- Shoemaker, P. J., & Vos, T. (2009). *Gatekeeping theory*. Routledge.

- Stewart, E. (2019, December 11). *How Mike Bloomberg made his billions: a computer system you've probably never seen*. Vox. <https://www.vox.com/2020-presidential-election/2019/12/11/21005008/michael-bloomberg-terminal-net-worth-2020>
- Tandoc, E. (2014). Journalism is twerking? How web analytics is changing the process of gatekeeping. *New Media & Society*, 16(4), 559–575. <https://doi.org/10.1177/1461444814530541>
- Tandoc, E. (2018). Five ways BuzzFeed is preserving (or transforming) the journalistic field. *Journalism*, 19(2), 200–216. <https://doi.org/10.1177/1464884917691785>
- Tandoc, E. C. (2019). Journalism at the periphery. *Media and Communication*, 7(4), 138–143. <https://doi.org/10.17645/mac.v7i4.2626>
- Tandoc, E., & Jenkins J. (2016). Out of bounds? How Gawker's outing a married man fits into the boundaries of journalism. *New Media & Society*, 20(2), 581–598. <https://doi.org/10.1177/1461444816665381>
- Tandoc, E., Lim, J.Y., & Wu, S. (2020). Man vs. machine? The impact of algorithm authorship on news credibility. *Digital Journalism*, 8(4), 548–562. <https://doi.org/10.1080/21670811.2020.1762102>
- Tandoc, E., & Oh, S.K. (2017). Small departures, big continuities? Norms, values, and routines in The Guardian's big data journalism. *Journalism Studies*, 18(8), 997–1015. <https://doi.org/10.1080/1461670X.2015.1104260>
- Tiffen, R., Jones, P. K., Rowe, D., Aalberg, T., Coen, S., Curran, J., Hayashi, K., Iyengar, S., Mazzoleni, G., Papatthanasopoulos, S., Rojas, H., & Soroka, S. (2014). Sources in the news. A comparative study. *Journalism Studies*, 15(4), 374–391. <https://doi.org/10.1080/1461670X.2013.831239>
- Van Dalen, A. (2012). The algorithms behind the headlines: How algorithm-written news redefines the core skills of human journalists. *Journalism practice*, 6(5-6), 648–658. <https://doi.org/10.1080/17512786.2012.667268>
- Waddell, T. F. (2018). A robot wrote this? How perceived machine authorship affects news credibility. *Digital journalism*, 6(2), 236–255. <https://doi.org/10.1080/21670811.2017.1384319>
- Wang, Q. (2018). Dimensional field theory: The adoption of audience metrics in the journalistic field and cross-field influences. *Digital Journalism*, 6(4), 472–491. <https://doi.org/10.1080/21670811.2017.1397526>
- Weaver, D. H., & Willnat, L. (2016). Changes in US journalism: How do journalists think about social media? *Journalism Practice*, 10(7), 844–855. <https://doi.org/10.1080/17512786.2016.1171162>
- Wu, S., Tandoc Jr, E. C., & Salmon, C. T. (2018). Journalism reconfigured: Assessing human-machine relations and the autonomous power of automation in news production. *Journalism Studies*, 20(10), 1440–1457. <https://doi.org/10.1080/1461670X.2018.1521299>
- Wu, S., Tandoc Jr, E. C., & Salmon, C. T. (2019a). When journalism and automation intersect: Assessing the influence of the technological field on contemporary newsrooms. *Journalism Practice*, 13(10), 1238–1254. <https://doi.org/10.1080/17512786.2019.1585198>
- Wu, S., Tandoc, E. C., & Salmon, C. T. (2019b). A field analysis of journalism in the automation age: Understanding journalistic transformations and struggles through structure and agency. *Digital Journalism*, 7(4), 428–446. <https://doi.org/10.1080/21670811.2019.1620112>
- Wu, Y. (2020). Is automated journalistic writing less biased? An experimental test of auto-written and human-written news stories. *Journalism Practice*, 14(8), 1008–1028. <https://doi.org/10.1080/17512786.2019.1682940>
- Zhang, S., & Feng, J. (2019). A step forward? Exploring the diffusion of data journalism as journalistic innovations in China. *Journalism Studies*, 20(9), 1281–1300. <https://doi.org/10.1080/1461670X.2018.1513814>

Biographical notes

Edson C. Tandoc Jr. is an Associate Professor at the Wee Kim Wee School of Communication and Information at Nanyang Technological University in Singapore. His research focuses on the sociology of message construction in the context of digital journalism. He has conducted studies on the construction of news and social media messages. His studies about influences on journalists have focused on the impact of journalistic roles, new technologies, and audience feedback on the various stages of the news gatekeeping process. For example, he has done some work on how journalists use web analytics in their news work and with what effects. This stream of research has led him to study journalism from the perspective of news consumers as well, investigating how readers make sense of critical incidents in journalism and take part in reconsidering journalistic norms; and how changing news consumption patterns facilitate the spread of fake news.

ORCID iD: 0000-0002-8740-9313

Scopus ID: 35751674400

Address: Wee Kim Wee School of Communication and Information, Nanyang Technological University, Room 02-39, 31 Nanyang Link, 637718, Singapore

Shangyuan Wu is a lecturer and media researcher at the Department of Communications and New Media, Media, Faculty of Arts and Social Sciences, National University of Singapore where she teaches media writing, journalism, communication management, and cultural studies. She graduated with a PhD in Communication from Simon Fraser University in Canada, where she also taught and researched media and communication for ten years before her return to Singapore. Her research areas of interest are centered on journalism in the digital age, with a focus on automated, data, immersive, and online journalism. Her research projects have involved investigations into the impact of social, political, economic and/or technological forces on the future of the journalism industry. She has published in peer-reviewed journals such as *Journalism Studies*, *Information, Communication and Society*, and *Digital Journalism*. She worked previously as a senior broadcast journalist and presenter at Mediacorp Radio, covering the areas of politics, defense and education.

ORCID iD: 0000-0003-3733-7988

Scopus ID: 57203895495

Address: Department of Communications and New Media, Faculty of Arts and Social Sciences, National University of Singapore, AS6, 11 Computing Drive, Singapore 117416, Singapore

Jessica Tan is a business journalism lecturer at Nanyang Technological University Singapore, where she also teaches online magazine and final year features projects. Lately, she has been exploring news media innovation in the classroom through the News Media Lab. Prior to NTU, she worked as a journalist, and her work has appeared in leading publications such as Dow Jones Newswires, *Forbes Asia* and *The Straits Times*. She also writes short stories, and her short story, *Dragon Girl*, was anthologized in *Twenty-two new Asian short stories* in 2016. She earned her post-graduate degree in journalism at the Medill School of Journalism, Northwestern University (2002), and holds a BA in History from the National University of Singapore (1999).

Address: Wee Kim Wee School of Communication and Information, Nanyang Technological University, 31 Nanyang Link, 637718, Singapore

Sofia Contreras-Yap is a PhD candidate at the Wee Kim Wee School of Communication and Information at Nanyang Technological University in Singapore. Her research focuses on the intersection of journalism, advertising, and social media and the impact of technological innovation on these fields. Current studies investigate native advertising, news in social media, and the state of journalism and advertising in Asia.

ORCID iD: 0000-0002-5733-6889

Address: Wee Kim Wee School of Communication and Information, Nanyang Technological University, 31 Nanyang Link, 637718, Singapore

How to cite

Tandoc Jr., E.C., Wu, S., Tan, J., & Contreras-Yap, S. (2022). What is (automated) news? A content analysis of algorithm-written news articles. *Revista Media & Jornalismo*, 22(41), PP. 103-120. DOI: https://doi.org/10.14195/2183-5462_41_6

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License