Full length article

Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists

CrossMark

Jaemin Jung [a], Haeyeop Song [a, *], Youngju Kim [b], Hyunsuk Im [a], Sewook Oh [b]

[a] School of Management Engineering, KAIST, Seoul, South Korea
[b] Korea Press Foundation, Press Center Bldg., Taepyeongno 1-ga, Jung-gu, Seoul, 100-750, South Korea

ABSTRACT

This study adopted a two (author: algorithm or journalist) by two (notification of author: real or inverse) between subject design to investigate how the public and journalists perceive the quality of algorithms-written articles compared with human journalist's work. Findings showed that both the public and journalists' evaluations were varied by the manipulation of author notification. That is, the public gave higher scores to the algorithm's work when it was notified as the real author, but they gave lower scores to the algorithm's work when the author was notified as a journalist. It confirmed the public's negative attitude toward journalists' credibility and craving for new information and communication technology (ICT) products/services in Korea. Based on journalists' resistance to change and innovation and the theory of prejudice, it was expected that journalists would be favorable to another journalist's work and unfavorable to an algorithm's work. However, contrary to the hypothetical expectation, journalists also gave higher scores to an algorithm's work and lower scores to a journalist's work. Implications relating to the intrusion of algorithm-written articles into journalism were discussed.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

The rise of the Internet has changed not only how news is distributed and consumed but also how it is produced. The Internet has drawn more people and organizations into news production. Beyond journalists in traditional news companies, there are also professional bloggers and online news startups. Even the general public acts as news providers. A recent development in the realm of journalism is software-generated content. Journalism entered a new phase with the rise of computer-written or automatically produced news articles. This advanced technological development has led to a new type of journalism—robot journalism.

News companies have started to work with algorithms to operate and publish software-generated news articles. Templates are produced in journalistic but iterative processes that require painstaking manual work. Although the technology is still in an early market phase, automated journalism has arrived in newsrooms. For example, *Forbes* uses an artificial-intelligence platform provided by the technology company *Narrative Science* to generate automated news on corporate earnings and stock prices from live datasets and content harvested from previous articles (Gani & Haddou, 2014). *AP* partnered with *Automated Insights* to begin automating quarterly earnings reports and now publishes 3000 such financial stories every quarter (Miller, 2015). After an earthquake hit Los Angeles one morning, it took only 3 min for the *LA Times* to write and publish an article about it online. It was written by an algorithm (Neal, 2014).

Beyond the U.S. news media, companies in other regions introduced algorithm-written articles. Chinese social and gaming giant Tencent published its first business report written by an algorithm in September 2015. It was written in Chinese and completed in just 1 min by Dreamwriter, a Tencent-designed algorithm (He, 2015). *The Financial News* in Korea ran an article reporting stock market news written by IamFNBOT, an algorithm writer (Son, 2016).

Algorithms are used not only in news writing but also in news selection and editing. In the *Guardian* from the U.K., the algorithmic newspaper software selects the most popular *Guardian* articles and

assembles them into a weekly printed newspaper (Ellis, 2013). It is well known that search engines like Google also use algorithms to display news on its site.

As the technologies are intruding into the creation of news, scholars have started to examine the changing nature of journalism amid data abundance, computational exploration, and algorithmic emphasis with growing significance for the media industry and for journalism as practice and profession (Lewis, 2015; also see the special issue: "Journalism in an era of big data" of *Digital Journalism*, 2015). Among the data-oriented practices emerging in journalism, Carlson (2015) noted that none appear to be as potentially disruptive as automated journalism, insofar as it calls up concerns about the future of journalistic labor, news compositional forms, and the very foundation of journalistic authority. Meanwhile, automated content with no human intervention beyond the initial programming can make journalists free from handling basic works and afford time to focus on more investigative reporting. It could be an opportunity for media companies to reinvent news production system by generating news faster, at a larger scale, and with fewer errors. It also could be beneficial for audience pursuing more news and unbiased reporting.

In fact, journalists are not among the first to feel both the pressure and opportunity of automation. There have been periodic warnings in the last two centuries that automation and new technology were going to wipe our large number of middle class jobs (Autor, 2015). Such concerns have recently regained prominence. In their widely discussed book *The Second Machine Age*, Brynjolfsson and McAfee (2014) offered an unsettling picture of the likely effects of automation on employment. They asserted that there's never been a better time to be a worker with special skills or the right education, because these people can use technology to create and capture value. However, there's never been a worse time to be a worker with only 'ordinary' skills and abilities to offer, because computers, robots, and other digital technologies are acquiring these skills and abilities at an extraordinary rate (Brynjolfsson & McAfee, 2014, p. 11).

In this new era of automated journalism, to activate discussion on automated news creation and its implications for journalists, journalism, news industry, and audience, it is necessary to explore the public's and journalists' perceptions and evaluation of news written by algorithms. How does the public perceive and evaluate the quality of computer-written news articles? Can they tell the differences between articles written by a human journalist and algorithms? Journalists perceive themselves as a professional group differentiated with the public, because they have expertise and duty. Then, how do journalists' perception and evaluations of algorithm-written articles compare with those from the public? This study compares the perceived quality of algorithm-written articles with the work by human journalists. To answer the questions, we sampled both general readers and journalists and conducted a series of experiments with human- and algorithm-written articles.

2. Literature review

The discourse around the use of computers and software to gather, distribute, and publish content has different kinds of labels (Clerwall, 2014). One term embracing it broadly is "computational journalism," which is described as "the combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism" (Hamilton & Turner, 2009, p. 2). Other terms are "robot journalism" (Clerwall, 2014; Dawson, 2010; Van Dalen, 2012) and "automated content" and "algorithmic news" (Anderson, 2013; Bunz, 2010; Levy, 2012).

Computational and algorithmic journalism might reshape the cultural practice of news creation due to the hybrid nature of newsroom sense-making technologies. Anderson (2013) argues that the human becomes partially obdurate and the material partly intentional. For this reason, there have been increasing attention to the impact of algorithm and data in journalism. Some of recent studies focused on the computational journalism using qualitative interviews and textual analysis. Young and Hermida (2015) found that computational thinking and techniques emerged in a (dis) continuous evolution of organizational norms, practices, content, identities, and technologies that interdependently led to new product. Carlson (2015) contended with the emergent practice of automated news content creation both in how it alters the working practices of journalists and how it affects larger understandings of what journalism is, and how it ought to operate.

Other studies examined the relationship between journalism and data. Anderson (2015) investigated historicized relationship between journalism and big data, and then asserted we need to consider the material objects (whether interviews, documents, human observations, or other objects) that underlie journalistic processes. Parasie (2015) examined the question about to what extent would the processing of huge datasets allow journalists to produce new types of revelation based on the epistemological approach.

Diakopoulos (2015) identified the algorithmic power as something worthy of scrutiny by computational journalists interested in accountability reporting. He tried to show how transparency might be used to effectively adhere to journalistic norms in the use of newsroom algorithms. Using algorithm as part of the journalistic process is not, by any means, a new phenomenon. However, the use of software to actually write a news story is a new advancement (Clerwall, 2014). Being rather new, less attention has been given to automatically generated articles and their implications for journalistic practice and audience response, especially in empirical studies.

With the advent of algorithms-written articles, scholars attempted to test the quality of the automatically produced article. The first study was conducted by Christer Clerwall (2014) in Sweden. He investigated how readers perceive software-generated content in relation to similar content written by a journalist. Although there were no statistically significant differences due to the limited sample size—46 respondents—the experiment revealed interesting and mixed results. While the software-generated content was perceived as more descriptive and boring, it was also considered to be more informative, objective, accurate, and trustworthy. In this study, the respondents were also asked to assess whether the text had been written by a journalist or by a computer. Of the 27 respondents who read the software-generated text, 10 thought a journalist wrote it and 17 thought it was software-generated. Out of 18 journalists, 8 thought a journalist wrote it, but 10 thought software wrote it. In other words, respondents failed to assess the author of the article correctly—whether it was written by a journalist or a computer. Overall, journalistic content produced by an algorithm was not or was barely discernible from content written by a journalist in this experiment.

Another study explored the perceived credibility of algorithm-written news articles, searching specifically for differences and similarities between journalists and news consumers in the Netherlands (Van der Kaa & Krahmer, 2014). In this study, 168 native Dutch speakers and 64 Dutch journalists were asked to evaluate the perceived levels of the expertise and trustworthiness of news articles (sports and finance news) written by algorithms. Similar to the previous study (Clerwall, 2014), news consumers perceived the levels of trustworthiness and expertise of the algorithm and journalist equally.