



<http://www.diva-portal.org>

Preprint

This is the submitted version of a paper published in *Journalism Practice*.

Citation for the original published paper (version of record):

Clerwall, C. (2014)

Enter the Robot Journalist: Users' perceptions of automated content.

Journalism Practice

<http://dx.doi.org/10.1080/17512786.2014.883116>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kau:diva-31596>

Enter the robot journalist: users' perceptions of automated content

Christer Clerwall, Ph.D.

Dept. of Geography, Media and communication

Karlstad University

Note to reader: This is the version first sent to Journalism Practice, i.e. **not the final version**. The final version is published at:

<http://www.tandfonline.com/doi/full/10.1080/17512786.2014.883116#.UxV3Ml4gd5g>

Abstract

The advent of new technologies has always spurred questions about changes in journalism – its content, its means of production, and its consumption. A quite recent development in the realm of digital journalism is software-generated content, i.e. automatically produced content. Companies such as Automated Insights offer services that, according to themselves “humanizes big data sets by spotting patterns, trends and key insights and describing those findings in plain English that is indistinguishable from that produced by a human writer” (Automated Insights, 2012).

This paper seeks to investigate how readers perceive software-generated content in relation to similar content written by a journalist. The study utilizes an experimental methodology where respondents were subjected to different news articles that were written by a journalist or software-generated. The respondents were then asked to answer questions about how they perceived the article; its overall quality, credibility, objectiveness etc.

The paper presents the results from a first small-scale study and they indicate that the software-generated content is perceived as, for example, descriptive, boring and objective, but not necessarily discernable from content written by journalists. The paper discusses the results of the study and its implication for journalism practice.

Keywords: robot journalism; automated content; experimental study; online journalism

Introduction

Our technology humanizes big data sets by spotting patterns, trends and key insights and describing those findings in plain English that is indistinguishable from that produced by a human writer.

(“Automated Insights - Products and Solutions,” 2012)

Imagine a car driving down a dark road. Suddenly a moose crosses the road. The driver fails to react in time, and the car crashes into the moose in high speed. The car, being equipped with a modern collision detection technology as well as a GPS, sends information about the collision to the appropriate authorities. At the same time, data about the accident is gathered by a news story service, and in a few seconds a short news story is written and distributed to subscribing online newspapers. At the online newspapers algorithms in the content management system (CMS) make the judgment that this is a story that will attract reader interest, forward it to the online editor, together with a recommendation of the positioning (e.g. “this is a top 10 story”), who finally approves the story for publishing.

This introductory example might seem a bit far-fetched. However, in light of developments in automated content production (exemplified in the quote above), I would like to argue that it is not.

Automated content can be seen as one branch of what is known as algorithmic news, others being adaption to SEO logics (Dick, 2011), click-stream logics (Karlsson & Clerwall, 2013), and search-engine based reporting – i.e. reporters being assigned to “stories” based on popular searches on for example Google, AOL Seed being one example). This type of algorithmic news is not concerned about what the public needs to know in order to make informed decisions and act as citizens in a democracy, but rather

what the public, at a given moment, seem to “want” (i.e. the public as consumers rather than as citizens)¹.

The advent of services for automated news stories raises many questions; e.g. what are the implications for journalism and journalistic practice, can journalists be taken out of the equation of journalism, how is this type of content regarded (in terms of credibility, overall quality, overall liking, to mention a few aspects) by the readers?

Scholars have previously studied, and discussed, the impact of technological development in areas such as how it affects, and/or is being adopted in, the newsrooms (e.g. Cottle & Ashton, 1999), journalism practice (Franklin, 2008; Pavlik, 2000), and how journalists relate to this development and their role as journalists (Van Dalen, 2012). van Dalen (2012) has studied how journalists relate to the development of automated content and their role/profession as journalist. However, to present date, the focus has been on “the journalists” and/or “the media”, and no one has investigated how the readers perceive automated content. Consequently, this paper presents a small-scale pilot study that seeks to investigate how readers perceive software-generated content in relation to similar content written by (human) journalists. The study draws on the following empirical research questions:

RQ1 – How is software-generated content perceived by readers, in regards to overall quality and credibility?

RQ2 – Is the software-generated content discernable from similar content written by human journalists?

Literature review

This section is divided into two parts. The first section briefly reviews previous research on various kinds of algorithmic, automated, and/or computational journalism. The second section presents research on assessment of journalistic content.

The discourse about the use of computers and software to gather, produce, distribute, and publish content, uses different kinds of labels. One such term is “computational

¹ For a brief introduction to automated content and the discussion about it, see (Bercovici, 2010; Bunz, 2010; Dawson, 2010; Levy, 2012).

journalism”, 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” (Dawson, 2010; van Dalen, 2012), “automated content”, “algorithmic news” (Anderson, 2012; Bunz, 2010; Levy, 2012) and the like.

Using technology as part of the journalistic process is not, by any means, a new phenomenon. The use of software to actually write journalistic content, however, is. Thus, being a rather new phenomenon, research about the use of automated content, and its implications for journalistic practice is quite scarce. However, recent research (e.g. Anderson, 2012; Flew, Spurgeon, Daniel, & Swift, 2012; van Dalen, 2012) has shed some light on this new phenomenon; how it may be studied and its implications for journalism.

Assessing content

“Quality” is a concept with many facets. For content, “quality” can refer to an overall, but somewhat vague, notion of “this was a really good article”, for example. “Quality” can also be assessed by different standards depending on who the receiver is. A professor in literature may value the linguistic structure of a journalistic text, while disregarding the poor use of sources, whereas a journalist may focus solely on the poor use of sources.

However many-faceted, when it comes to assessing “quality” in news reporting, one concept stands out as the most important one, and that is *credibility*. In past studies, when users are asked to evaluate the quality of, for example, online news, it is actually the credibility that is (usually) being assessed. Hence, the literature review below focuses on studies on credibility in information (online) in general and news (online and elsewhere) in particular.

Previous research shows that credibility, as a concept, can be, and has been, studied along various dimensions such as, for example, source credibility, message credibility, and media credibility (Chung, Kim, & Kim, 2010, p. 672). Since the first (as acknowledged by for example Chung, Kim, & Kim, 2010; Sundar & Nass, 2001) study on credibility, by Hovland and Weiss (1951), a plethora of studies have been undertaken on the matter. Flanagin & Metzger (2000, 2007) have studied what affects users’

perception of credibility of websites and what type of “information verification procedures” (Flanagin & Metzger, 2000, p. 518) they use to assess the credibility of the site content. On a similar track, Fogg et al. (2003) focus on what users of websites consider to be important factors when assessing the credibility of websites. The results from their study show that when users are asked to comment on credibility issues, “design look” is in the top, with “information design/structure” and “information focus” as first and second runners up. (Fogg et al., 2003) Their study is on websites in general, but when the data is broken down they notice that for news sites, “information bias” is mentioned more than the overall average (appr. 30% compared to appr. 12%).

In a recent study Westerwick (2013) has studies how the credibility of online information is affected by the message itself (e.g. information quality, accuracy, currency, and language intensity), the “sponsor” of the web sites (e.g. “evaluations of the Web site’s sponsor, which may result from expertise or personal experience with the organization, group or person” (pp. 195-196), and the web site in itself (e.g. formal site features such as “visual elements, the amount of information provided on the site, or the degree of interactivity offered to the visitor” (p. 196)).

Newhagen and Nass (1989) study users’ criteria for evaluating credibility of news in newspapers and in television. One important contribution from their study is the idea that users use different sets of criteria for assessment depending on the medium where the news is presented; for newspapers the users focus on criteria pertaining to the newspaper as an institution (i.e. do they have confidence in the specific newspaper at large), while for television news they focus on “an aggregate of on-camera personalities” (Newhagen & Nass, 1989, p. 284). Their findings have implications for the research design of studies on credibility assessment.

In a similar way, Chung et al. (2010) studies what they call the anatomy of credibility in a study of online news. From a set of variables they yield three overall components (based on factor analysis) pertaining to credibility; *expertise*, *trustworthiness*, and *attractiveness*. Although their study’s main focus is on comparing different types of news sites (mainstream, independent and indexes), the scales used, as well as the resulting components are of interest for the present study.

In the process of assessment Flanagin & Metzger (2000) noticed that users make use of previous knowledge about a certain genre and then interpret the information in accordance to that knowledge. As readers of mass media, users know that the information have passed through a “filter” that should provide for a degree of information reliability, i.e. the information should be factually correct and (at least somewhat) objective.

The research on news content has been focused on credibility, but there are other aspects that might be of interest in the evaluation of news content. Sundar (1999) notices this and subsequently the first step in his study is a pretest where respondents were asked to read news articles and then list the thoughts that came to their mind (with a focus on adjectives related to the articles). As noted by Sundar, this is a useful way to get the view of the respondents that is not colored by the researches preconceptions.

One of the contributions from the Sundar (1999) study is a set of factors (or components) resulting from a factor analysis on 21 measures. These factors are:

- **Credibility**; described as “...a global evaluation of the objectivity of the story”,
- **Liking**; described as “...overall affective reaction”, and as “an indicator of a news receiver’s feelings toward [...] the overall content of the news story”,
- **Quality**; as “the degree or level of overall excellence of a news story” (informative, important, interesting, well-written).
- **Representativeness**; described as “a summary judgment of the extent to which the story is representative of the category of news”. (Sundar, 1999, p. 380f)

So, again, *credibility* seems to be at the heart of assessment of news stories. However, as list above indicate, there are other criteria pertaining to the evaluation of the *overall quality* of a news story. For example, Slater and Rouner (1996) studies message quality evaluation and note that the text organization and style (“consistency of tone, uniqueness of voice, presence of attitude [...], level of formality, creativity, and more”) influences the message evaluation, and that a “[f]avorable evaluation of a message may mean it is perceived as well written, and it brings the reader closer emotionally and cognitively” (Slater & Rouner, 1996, p. 976)

Based on the literature review above, Table 1 lists concepts/descriptors with respect to content evaluation.

Table 1. Descriptors of credibility and quality in content

Descriptor		Suggested by
Factual		(Newhagen & Nass, 1989)
Can be trusted/believable		(Newhagen & Nass, 1989; Sundar, 1999)
Fair		(Chung et al., 2010; Newhagen & Nass, 1989; Sundar, 1999)
Accurate		(Chung et al., 2010; Newhagen & Nass, 1989; Sundar, 1999)
Tell the whole story/comprehensive/in-depth		(Chung et al., 2010; Newhagen & Nass, 1989; Sundar, 1999)
Reporters are well trained/ Written by professional journalists		(Chung et al., 2010; Newhagen & Nass, 1989)
Separated facts from opinions		(Newhagen & Nass, 1989)
Concerned about the community's well being/Working for the public good		(Chung et al., 2010; Newhagen & Nass, 1989)
Unbiased/Biased		(Chung et al., 2010; Newhagen & Nass, 1989; Sundar, 1999)
Moral		(Newhagen & Nass, 1989)
Watches out after your interests		(Newhagen & Nass, 1989)
Sensationalizes/sensationalistic		(Newhagen & Nass, 1989; Sundar, 1999)
Respects people's privacy		(Newhagen & Nass, 1989)
Patriotic		(Newhagen & Nass, 1989)
Does not care what the reader thinks		(Newhagen & Nass, 1989)
Objective		(Chung et al., 2010; Sundar, 1999)
Boring		(Sundar, 1999)
Enjoyable		(Sundar, 1999)
Interesting		(Sundar, 1999)
Lively		(Sundar, 1999)
Pleasing/attractive		(Chung et al., 2010; Sundar, 1999)
Clear		(Sundar, 1999)
Coherent		(Sundar, 1999)
Concise		(Sundar, 1999)
Well-written		(Sundar, 1999)
Disturbing		(Sundar, 1999)
Important		(Sundar, 1999)
Relevant/useful		(Chung et al., 2010; Sundar, 1999)
Timely		(Sundar, 1999)

Informative		(Chung et al., 2010; Sundar, 1999)
Professional		(Chung et al., 2010)
Delivering a diversity of opinions		(Chung et al., 2010)
Authoritative		(Chung et al., 2010)
Creative		(Chung et al., 2010)
Interest		(Chung et al., 2010)
Well-written		(Slater & Rouner, 1996)
Consistency of tone		(Slater & Rouner, 1996)
Creative		(Slater & Rouner, 1996)
Level of formality		(Slater & Rouner, 1996)

Although most of descriptors in Table 1 pertain to the message, some of them can also be applied to the source. The message in itself is one thing, but Slater and Rouner also emphasize the importance of “source”, or where the message originates from: “It seems self evident that if a message originates with, for example, an expert and objective person, that message should influence audience beliefs more than the same message from an inexperienced and biased person. Presumably, the audience member is cued by that source attribution to employ different processing strategies that result in the subsequent message arguments' being more readily accepted or rejected.” (Slater & Rouner, 1996, p. 975)

So, to summarize this section, as noted by previous research (e.g. Chung et al., 2010; Flanagin & Metzger, 2000, 2007; Hovland & Weiss, 1951; Newhagen & Nass, 1989; Sundar, 1999) the medium and/or channel as well as the source/sender is important in the users' perception of the credibility of a message. However, when a message is stripped of these credibility “clues”, users will have to draw conclusions about the message based on how they perceive the quality of the message: This is affected by the *presentation*, *plausibility* and the *specificity* of the message (Slater & Rouner, 1996).

Following Slater & Rouner, the focus of this study is on the message itself, and in the following section the study and its experimental setup is described.

Method

The focus of this study is on the *text* or *message*. In order to be able to study how a piece of software-generated content is perceived, in relation to a similar text, written by a journalist, the study facilitated an experimental research method.

The research design for the study comprised three stages – a pre-test, a small test of the survey, and the pilot-test.

As Sundar (1999) has noted, "...the values obtained by researchers on particular measures are at least in part a function of the fact that the researchers elicited participants' values, and not necessarily an indication of the relevant psychological dimension(s) along which participants vary in response to stimuli." (p. 374). To address this issue, a pre-test with students attending a research methodology class in media and communication studies was conducted with the aim to obtain the respondents unbiased (in the sense that they are not colored by the preconceptions of the researcher) evaluations of the news stories, and to obtain descriptors not found in the table above. The respondents were asked to read a text and were then encouraged to assign at least five words or phrases to the article. The pre-test did not yield any new descriptors and so the descriptors in the table above were deemed as sufficient for the following test.

For the small test of the survey form, seventeen students in a web production course were randomly assigned to one of two treatments. They were asked to read a game recap (either written by a journalist or software-generated), and were then asked to assess the text on twelve variables. The test also included the possibility to add other descriptors (once again to see if there were descriptors that had been overlooked). The test worked well, no new descriptors were found and the larger pilot-test could be launched.

In the pilot-test, undergraduate students in media and communication studies were invited to take part in the experiment. The students were randomly assigned to each treatment and an invitation was sent via e-mail. In total, 46 students participated in the test, 30 women (65 percent) and 16 men (35 percent). The youngest participant was 20 years old and the oldest was 32. The mean age was 23. 19 respondents were assigned to group 1 (journalist) and 27 were assigned to group 2 (software).

For the test, a software-generated game recap (from the Bolts Beat web site, Appendix 1) was collected and trimmed to just include text. For the text written by a journalist, an article on NFL (from latimes.com, Appendix 2) was collected. The text was shortened to match the length of the game recap. The test facilitated a web survey where the participants were exposed to either the text written by a journalist (group 1) or the software-generated game recap (group 2).

After reading the text, the participants were asked to assess the article on the quality of the content, and the credibility of the text. The twelve descriptors used were derived from the Table 1 above and they are (as they appeared in the survey): *objective, trustworthy, accurate, boring, interesting, pleasant to read, clear, informative, well written, useable, descriptive, and coherent*.

After making this assessment, the respondents were also asked to assess whether they thought the text was written by a journalist or generated by software.

Results

To answer the first RQ – How is software-generated content perceived by readers, in regards to overall quality and credibility? – the respondents were asked to assess the text on twelve descriptors (accounted for above). For the twelve descriptors the respondents were asked how well the descriptor fitted the text: “Below you will find words that can describe the text above. Please assess to which degree you think the text was...”. This was followed by the twelve descriptors in a matrix setup with a five-point Likert-scale for each descriptor (ranging from 1 *Not at all* to 5 *Totally*), see Appendix 1 for a screenshot of the setup.

Although the respondents were randomly assigned to each treatment, the Likert-scale does not meet the requirements for parametric tests (for a brief discussion on this matter see Field & Hole (2010)). Thus, in order to compare the groups, mean ranks were calculated using a Mann-Whitney test for each descriptor. Figure 1 presents the mean ranks for each group:

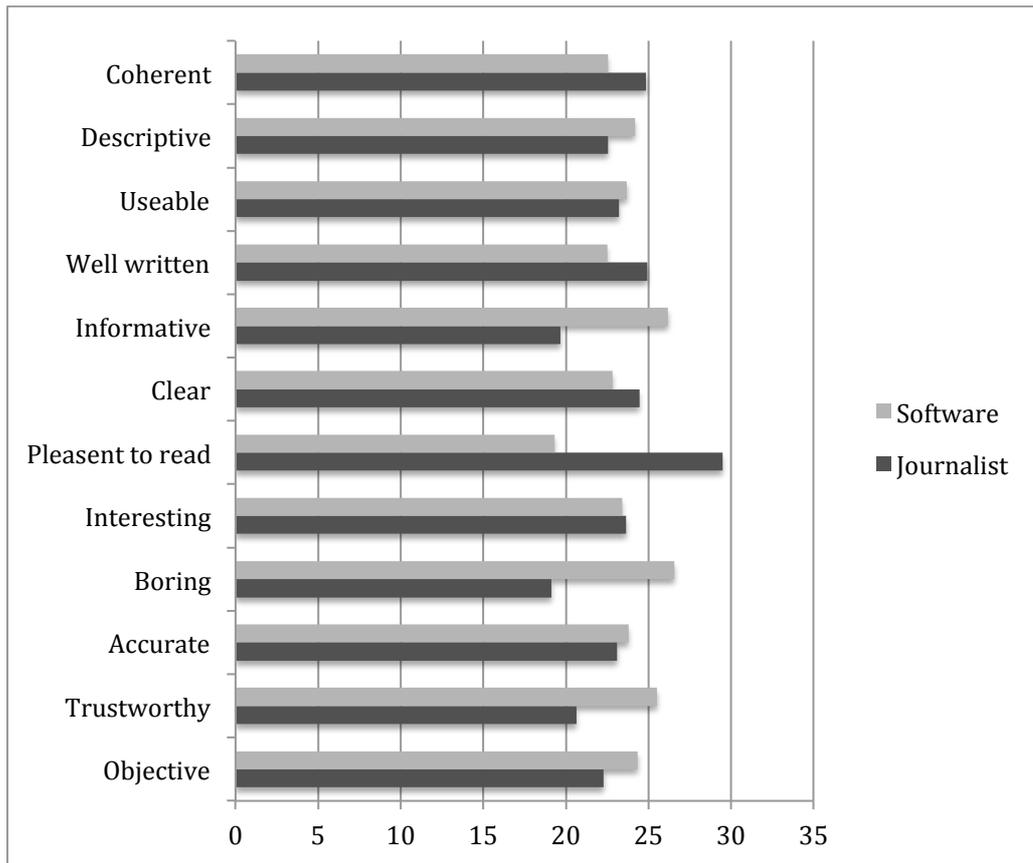


Figure 1 - Mean rank values for each descriptor, clutred by group (journalist or software)

Looking at the bars in Figure 1, we can see that the text written by a journalist scores higher on *coherence*, “*well written*”, “*clear*”, and on being *pleasant to read*. Since *boring* is a negative judgment the journalist written text can be said to score better on this as well. However, the software-generated text scores higher on other descriptors, such as *being descriptive* (whether or this is a positive may of course be a matter of personal preferences), *informative*, *trustworthy*, and *objective*.

Although the differences is small, the software-generated content can be said to score higher on descriptors typically pertaining to the notion for credibility.

To further compare the values, Figure 2 below shows a balance-score for each descriptor. The score is calculated by subtracting the mean rank value of the software-generated text from that of the text written by a journalist.

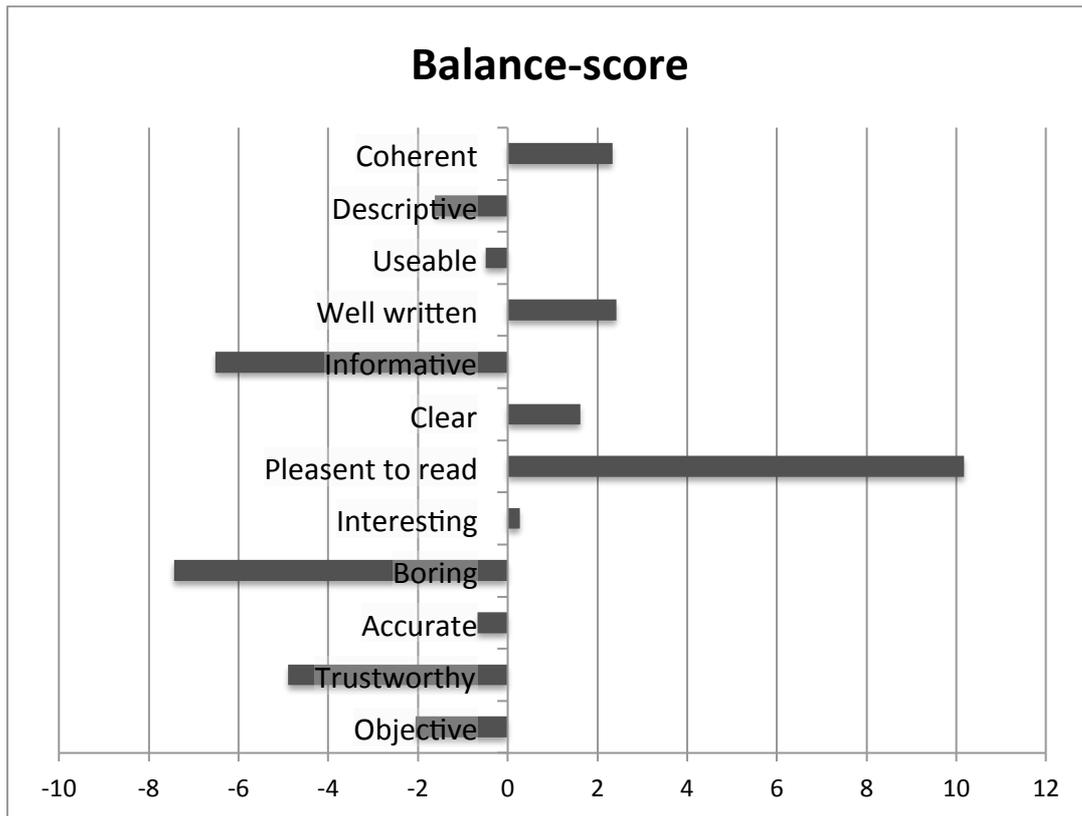


Figure 2 - Illustrates the balance-score for each descriptor. To reach the balance-score, the mean rank value of the software-generated text was subtracted from the mean rank value of the text written by a journalist.

Based on Figure 2, we can say that the text written by a journalist is assessed as being more coherent, well written, clear, less boring and more pleasant to read. On the other hand, the text generated by a software is perceived as more descriptive, more informative, more boring, but also more accurate, trustworthy and objective.

But are these differences significant? The short answer is, no they are not. Using a Mann-Whitney test for a non-parametrical independent samples, only the descriptor *Pleasant to read* showed a statistically significant ($U=143$, $r=-0.39$) difference between the two treatments. The lack of significant differences will be discussed further in the discussion section of this paper.

The second research question for the study was *Is the software-generated content discernable from similar content written by human journalists?* In the second part of the survey, the respondents were asked to assess whether or not the text had been written by a journalist or by a computer (software).

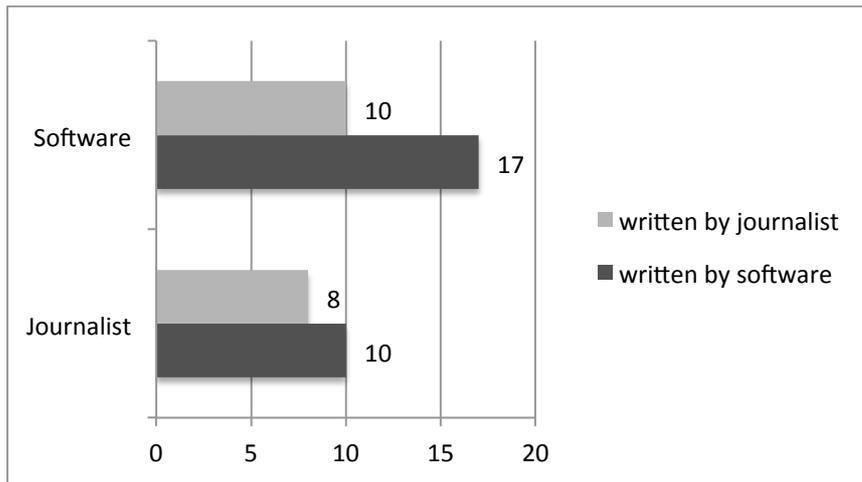


Figure 3 - Respondents assessment about the origin of the text (software or journalist). N=45 (one answer missing)

Of the 27 respondents who read the software-generated text, 10 thought a journalist wrote it, and 17 thought it was software-generated. For the 18 respondents in the “Journalist group”, 8 perceived it as having been written by a journalist, but 10 thought a software wrote it.

Using a Mann-Whitney test for significance we can conclude that there is no significant difference ($U=225$, $r=-0.07$, $\text{sig.}=0.623$) between how the groups have perceived the texts.

Discussion

Perhaps the most interesting result in the study is that there are no (with one exception) significant differences in how the two texts are perceived by the respondents. The lack of difference may be seen as an indicator that the software is doing a good job, or it may indicate that the journalist is doing a poor job – or perhaps both are doing a good (or poor) job?

If journalistic content produced by a piece of software is not (or is barely) discernable from content produced by a journalist, and/or if it is just a bit more boring and pleasant to read, then why should news organizations allocate resources to human writers?

Perhaps the speed, an important factor in adopting new technologies, (cf. Örnebring, 2010), will make up for the loss of “pleasantness”? If the audience can get automated content cheaper than content produced by journalists, with “less pleasant to read” as the main drawback – why would they want to pay?

As Pavlik (2000, p. 229) has noted “[j]ournalism has always been shaped by technology”. This is not say that technology in and of itself drives change, but it is an intrinsic part of the mix of economical, political, social, organizational, to mention a few, factors (Boczkowski, 2004; Örnebring, 2010) When it comes to automated content, as one technological factor amongst many, van Dalen (2012) notes that journalists recognizes that automated content may be a threat to some journalists as it “...may put journalists doing routine tasks out of work” and it “...can be applied beyond sports reporting and also challenge the jobs of journalists in finance or real estate” (Van Dalen, 2012, p. 8). In the same study the journalists emphasize a couple of strengths of human journalists as *creativity*, *flexibility* and *analytical skills* indicating that the more advanced journalism is not threatened by automated content.

As far as this study goes, the readers are not able to discern automated content from content written by a human. Some aspects of quality, such as *being clear* and *being pleasant to read* got a bit higher score for human written content, but others, such as *trustworthiness*, *informative*, and *objective* were higher for the automated content. However, how automated content may influence journalism and the practice of journalism is a quite open question. An optimistic view would be that automated content will free resources that will allow reporters to focus on more qualified assignments, leaving the descriptive “recaps” to the software. This type of positive outlook is purveyed by Flew et al. (2012, p. 167) “Ultimately the utility value of computational journalism comes when it frees journalists from the low-level work of discovering and obtaining facts, thereby enabling greater focus on the verification, explanation and communication of news.”

However, making use of automated content may just as well be seen as a way for news corporations to save money on staff, as they do not need the reporters to produce the content.

Limitations and further research

The result of this study is based on a small and quite skewed sample. A larger sample, representing a larger part of the general audience may yield a different result.

Furthermore, the articles used may not be very representative for neither software-

generated content, nor content written by a human. The experiment comprised **one** article from each category, making the risk for a skewed result quite apparent.

However, the result of the study makes a case for further research in order to more thoroughly study how the audience perceives automated content, how it is used by the news industry, and how it may impact the practice of journalism.

References

- Anderson, C. W. (2012). Towards a sociology of computational and algorithmic journalism. *New Media & Society*, (December).
- Automated Insights - Products and Solutions. (2012). Retrieved September 25, 2012, from http://automatedinsights.com/products_and_solutions#prod-sol-stories-articles
- Bercovici, J. (2010). Can You Tell a Robot Wrote This? Does It Matter? *Forbes*. Retrieved May 24, 2013, from <http://www.forbes.com/sites/jeffbercovici/2010/11/29/can-you-tell-a-robot-wrote-this-does-it-matter/>
- Boczkowski, P. J. (2004). The Processes of Adopting Multimedia and Interactivity in Three Online Newsrooms. *Journal of Communication*, 54(2), 197–213.
- Bunz, M. (2010). In the US, algorithms are already reporting the news. *Guardian*. Retrieved May 24, 2013, from <http://www.guardian.co.uk/media/pda/2010/mar/30/digital-media-algorithms-reporting-journalism>
- Chung, C. J., Kim, H., & Kim, J. H. (2010). An anatomy of the credibility of online newspapers. *Online Information Review*, 34(5), 669–685.
- Cottle, S., & Ashton, M. (1999). From BBC Newsroom to BBC Newscentre : On Changing Technology and Journalist Practices. *Convergence: The International Journal of Research into New Media Technologies*, 5(3), 22–43.
- Dawson, R. (2010). The rise of robot journalists. *Trends in the Living Networks*. Retrieved May 24, 2013, from http://rossdawsonblog.com/weblog/archives/2010/04/the_rise_of_rob.html
- Dick, M. (2011). SEARCH ENGINE OPTIMISATION IN UK NEWS PRODUCTION. *Journalism Practice*, 5(4), 462–477.
- Field, A., & Hole, G. (2010). *How to design and report experiments*. London: SAGE.

- Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet Information Credibility. *Journalism & Mass Communication Quarterly*, 77(3), 515–540.
- Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society*, 9(2), 319–342.
- Flew, T., Spurgeon, C., Daniel, A., & Swift, A. (2012). THE PROMISE OF COMPUTATIONAL JOURNALISM. *Journalism Practice*, 6(2), 157–171.
- Fogg, B. J., Soohoo, C., Danielson, D. R., Tauber, E. R., Stanford, J., & Marable, L. (2003). How Do Users Evaluate the Credibility of Web Sites? A Study with Over. *Proceedings of the 2003 Conference on Designing for User Experiences*.
- Franklin, B. (2008). The future of newspapers. *Journalism Practice*, 2(3), 306–317.
- Hamilton, J. T., & Turner, F. (2009). *Accountability Through Algorithm: Developing the Field of Computational Journalism*. Retrieved from <http://dewitt.sanford.duke.edu/wp-content/uploads/2011/12/About-3-Research-B-cj-1-finalreport.pdf>
- Hovland, C. I., & Weiss, W. (1951). The Influence of Source on Communication Credibility Effectiveness. *Public Opinion Quarterly*, 15(4), 635–650.
- Karlsson, M., & Clerwall, C. (2013). Negotiating Professional News Judgment and “Clicks”: Comparing tabloid, broadsheet and public service traditions in Sweden. *Nordicom Review*.
- Levy, S. (2012). Can machines write better news stories than humans? *Wired*. Retrieved June 5, 2013, from <http://www.wired.com/gadgetlab/2012/04/can-an-algorithm-write-a-better-news-story-than-a-human-reporter/>
- Newhagen, J., & Nass, C. (1989). Differential Criteria for Evaluating Credibility of Newspapers and TV News. *Journalism Quarterly*, 66(summer), 277–284. Retrieved from <http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=EJ398777>
- Pavlik, J. (2000). The Impact of Technology on Journalism. *Journalism Studies*, (1), 2.

- Slater, M. D., & Rouner, D. (1996). How message evaluation and source attributes may influence credibility assessment and belief change. *Journalism & Mass Communication Quarterly*, 73(4), 974–991.
- Sundar, S. S. (1999). Exploring Receivers' Criteria for Perception of Print and Online News. *Journalism & Mass Communication Quarterly*, 76(2), 373–386.
- Sundar, S. S., & Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication*, 51(1), 52–72.
- Van Dalen, A. (2012). The Algorithms Behind the Headlines. *Journalism Practice*, (September), 1–11.
- Westerwick, A. (2013). Effects of Sponsorship, Web Site Design, and Google Ranking on the Credibility of Online Information. *Journal of Computer-Mediated Communication*, 18(2), 80–97.
- Örnebring, H. (2010). Technology and journalism-as-labour: Historical perspectives. *Journalism*, 11(1), 57–74.

Appendix 1

The screenshot shows the setup for the first part of the test. It also shows the text for the game recap written by the Statsheet software.

Chargers Take Down Chiefs, 37-20

Even with an unexceptional outing for Philip Rivers, the Chargers handled the Chiefs, 37-20, at Arrowhead Stadium.

Rivers found the end zone for two touchdowns against the Chiefs on 18 of 23 passing for 209 yards and one pick. Matt Cassel went 24 of 42 with 251 yards passing, two touchdowns and three picks for the Chiefs.

Jackie Battle carried the ball 15 times for 39 yards with a touchdown in addition to four receptions for 42 yards and another touchdown. Antonio Gates caught three passes for 59 yards.

The San Diego defensive unit led the way to a victory, allowing 119 yards rushing and 234 passing while bringing back one interception for a touchdown. They brought down Cassel for two sacks.

Lastly, Nick Novak was perfect, hitting all three of the field goals he attempted.

Chiefs running back Jamaal Charles provided some spark with 92 yards rushing, 23 receiving and two touchdowns. Wide receiver Dwayne Bowe also gave a strong effort with 108 receiving yards and one touchdown.

The Chargers travel to New Orleans next week to meet the Saints (0-3) while the Chiefs are at home to take on the Ravens (3-1).

1. Nedan följer ord som kan beskriva texten ovan. Angetill vilken grad du tycker att texten var...

	Stämmer helt	Stämmer	Varken eller	Stämmer delvis	Stämmer inte alls
objektiv	<input type="radio"/>				
trovärdig	<input type="radio"/>				
korrekt	<input type="radio"/>				
tråkig	<input type="radio"/>				
intressant	<input type="radio"/>				
trevlig att läsa	<input type="radio"/>				
tydlig	<input type="radio"/>				
informativ	<input type="radio"/>				
välskriven	<input type="radio"/>				
användbar	<input type="radio"/>				
beskrivande	<input type="radio"/>				
sammanhängande	<input type="radio"/>				

Appendix 2 - Text written by journalist (shortened)

Three quarterbacks are walking a tightrope

Matt Cassel, Russell Wilson and Mark Sanchez have struggled, and their starting jobs are in jeopardy.

Their passes might sail high, but three NFL quarterbacks have landed far short of expectations.

Kansas City's Matt Cassel, Seattle's Russell Wilson, and the New York Jets' Mark Sanchez aren't the only starting quarterbacks who are struggling — there are several — but they're the ones inching ever closer to the bench.

Through four games, the three have combined for 14 touchdowns and 15 interceptions, and each plays for a team in danger of falling behind early in their respective division races.

In the brightest spotlight is Sanchez, and not only because he plays in the country's biggest market. He has Tim Tebow looking over his shoulder, and it's only a matter of time until the Jets give Tebow a chance — a telegraphed pass if there ever was one.