universität wien

MAGISTERARBEIT / MASTER'S THESIS

Titel der Magisterarbeit / Title of the Master's Thesis

„The Influence of the Topic on the different perception of computer generated versus human written news"

verfasst von / submitted by

Mathias Wagner, Bakk. BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Magister der Philosophie (Mag. Phil.)

Wien, 2019 / Vienna 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 841
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Publizistik- und Kommunikationswissenschaft
Betreut von / Supervisor:	Univ.-Prof. Dr. Homero Gil de Zúñiga, PhD

Table of Contents

1 Introduction

In today's media, there is a huge overflow of similar news produced by human journalists. In May 2019 a search on Google News with the keywords "Trump" and "Gun Control" found over eight million results in less than 0.5 seconds. Each one of those articles seems rather meaningless given the high number of stories on the same exact topic. It only makes sense that publishers think about saving costs by making their production more efficient in such an environment. Shoshana Zuboff's first "law" says: "Everything that can be automated will be automated." (Linden, 2017b)

And that is exactly what is happening in newsrooms around the world today. Algorithms that produce news content are used by an increasing number of news companies (Liu & Wei, 2018). Digital technologies like those algorithms are a threat to the survival of traditional news organizations. They alter the way news are gathered, distributed and consumed. They also contribute to the creation of an unprecedented number of news items (Wu, Tandoc & Salmon, 2019). Those news stories have been labeled as "Automated News", "Computer-Generated News", "Algorithmic News" and as "Robot News" written by "News Robots" (Liu & Wei, 2018). All those words are used as synonyms to describe the same exact phenomenon. The term "Robot" is a bit misleading in this context. The news items are generated by algorithms running on computers and not by physical robots. In this paper the author will use the words "Automated News" and "Computer-written News" (in distinction from Human-written news).

The main goal of this paper is to further improve our knowledge about the different perceptions of computer-written and human-written news. In the following chapter, the current literature on computer-written news will be reviewed. Computer-written news have been researched from a lot of different points of view in the past. The author will provide an

overview of the knowledge we have about this rather new technology. Previous work on the different perceptions of computer-written and human-written news will be discussed towards the end of the chapter and lead to the research questions and hypothesis of this paper. In chapter 3, the method used in this paper is described. The results are then provided in chapter 4 and discussed in chapter 5. Chapter 6 talks about the limitations the author had to deal with. Chapter 7 concludes and provides some ideas on future work that needs to be done in this field.

2 Literature Review

2.1 History of the usage

Automated News were first introduced in the United States in 2010 and a few years later they were also introduced in Europe. A good example of when Automated News gained popularity among producers is the departmental election in France in 2015. An algorithm generated over 34.000 pieces about the election for *Le Monde*. Ever since then news companies in France have woken up to the possibility of using Computer-Generated News. Another example of how Automated News are used in Europe is the Swedish Newspaper *Mittmedia*. It uses Computer-Generated News to produce local weather reports (Linden, 2017b).

More examples of how Automated News are used are described by Linden, 2017a. The *Associated Press* use algorithms to generate financial articles whenever earnings reports are released. Before the adoption of Automated News by the *AP*, human writers had to work a lot of hours to sort through the data and write a lot of articles that were all very similar to each other. Now an Algorithm does that. That way, the *AP* were able to increase their output from 300 financial stories on earning reports to over 3.700. Those articles contain fewer spelling and calculating mistakes than human-written articles. Another example mentioned by Linden, 2017a, is Local Labs. They use algorithms to write automated e-mails to people and

organizations, so that they know in advance when newsworthy events will take place. Local Labs argue that being proactive rather than reactive saves a lot of time and money.

2.2 Algorithms in news process

Algorithms can be used in different parts of the news production process. Possibilities include information gathering, organization of information, sense-making, and the actual communication and presentation themselves (Coddington, 2015). More precisely, algorithms can issue alerts to trending stories, collect and visualize data, auto-publish news and write news stories (Wu, Tandoc & Salmon, 2019).

The best-known algorithms in the entire process are probably news selection algorithms. The Brexit vote and the election of Donald Trump as President of the United States both led to public discussions about algorithms that select news on social media, particularly Facebook, and how those algorithms take over the role of human journalists as gatekeepers. This even led to discussions in traditional news outlets like the *BBC*, who try to achieve a balance between the different ways to decide which news items to present to users. The different possibilities include letting the audience themselves pick their favorite topics, using algorithms and having a common set of stories selected by journalists and editors (Thurman, Moeller, Helberger, & Trilling, 2019). The four authors conducted an international survey in 26 countries to find out which type of news selection the participants liked best. Their results show that algorithmic news selection based on the consumers past activities was rated best, especially among young people. As trust in news organizations goes down so does the trust in news selection by journalists, which is a remarkable result as most other papers on recommendations in any other field than news selection show, that participants usually prefer recommendations by human experts over recommendations by algorithms (ibid.).

The possibilities to use algorithms are not limited to getting information, generating articles and selecting news, obviously. Other examples of how algorithms are used in media companies are so called “News Bots” that operate on Social Networks (Lokot and Diakopoulos, 2016). Bots on Social Media or Bots that chat with users of products are probably best known in today’s world. Besides that, there are more possibilities to use algorithms now and there will be a lot more in the future. An interesting piece of work is the article by Gynnild, 2014, on Robot eyewitnesses. She describes how drones can be used in journalism to either supplement or replace the current methods of covering news events visually.

But since this paper focuses on Automated News, those technologies are just mentioned here for the sake of completeness. From now on this paper focuses entirely on Automated News.

2.3 Technical perspective

In this chapter the technical perspective of Automated News is summarized. Automated News are an example of Natural Language Processing (NLP). A definition of NLP is offered by Khurana, Koli, Khatter, & Singh: “Natural Language Processing […] is a tract of Artificial Intelligence and Linguistics, devoted to make computers understand the statements or words written in human languages.” (p. 1) The idea of humans talking to computers in their natural language is rather old and has long been part of science fiction movies. This science fiction dream was realized in the past decade by voice assistants like Apple’s Siri, Microsoft’s Cortana, Amazon’s Alexa and Google’s assistant, which are now available to a high number of people (Hoy, 2018).

Obviously, Automated News are a different application of NLP, but the idea is the same. Humans and computers communicate with each other in a language natural to humans. NLP can be divided into two parts, Natural Language Understanding and Natural Language Generation. The voice assistants, that were mentioned in the previous paragraph, are an

example of Natural Language Understanding. Automated News are an example of Natural Language Generation, obviously (Khurana, Koli, Khatter, & Singh).

The main difference between Natural Language Understanding and Natural Language Processing is that the algorithm must choose between different options when generating texts. Quickly summarized, texts are usually generated in three steps: In the first stage, called "Document Planning" (Reiter, 2010, p. 576), the algorithm decides what information is relevant. Obviously, the input data plays a large role here. The second stage is called "Microplanning" (ibid., p. 577). In this stage, the algorithm decides which linguistic structures (words, syntax, sentences) are going to be used. In the third and final stage, the "Realization" (ibid.), a text is generated based on the decisions of the first two stages. The linguistic structures are brought into an order. The product of that stage is a finished text (Reiter, 2010).

Obviously, Automated News have their limitations and can't be used to replace journalists entirely. At this point there is no indication that the algorithms will ever be able to do that. In order to perform their tasks those algorithms need a lot of data to analyze. They draw their conclusions and identify interesting facts and generate texts automatically, much faster than humans would be able to. The need for a high amount of data is one of the main limitations of News Algorithms. They are mostly used to generate sports and financial articles and weather and traffic reports (Dörr, 2016). Recently news companies have experimented with celebrity news, where large databases of celebrity relationships, appearances, awards and endorsements are stored. To sum up, Automated News are best used when a lot of stories need to be produced and when a lot of data is there for the algorithms to analyze. Once Automated News are generated, they are either published or used as initial drafts for journalists to save them time they would normally spend doing routine tasks like analyzing data or writing simple game reports (Haim & Graefe, 2017).

2.4 Motivation for using algorithms

It was already mentioned that more and more news companies produce Automated News. Obviously, there is a variety of reasons why executives make the decisions to do that. The main reason according to Kim & Kim, 2017, is market forecasting. Executives are highly sensitive to the outlook of whether news customers are willing to accept Automated News. Journalists' attitude towards Automated News are also important. Sunken costs are behind market forecasting, customers willingness and journalists' attitude on the list of the main reasons why publishers use or try using algorithms and not as important as one would think. It shows that the expected effects of the adoption of Automated News are more important at this point (Kim & Kim, 2017a). While there might be some other motivations to use algorithms to produce news right now, there is no doubt that money will be the ultimate reason why Automated News are here to stay. We have seen a massive decline in advertising revenue and circulation of newspapers in recent years which forces traditional news companies to adopt cost-cutting measures (Wu, Tandoc, & Salmon, 2019). Today, there are more different devices and platforms on which people consume news than ever, which implies that revenue models of news providers are more diverse than ever (Thurman, Moeller, Helberger, & Trilling, 2019).

Wu, Tandoc & Salmon, 2019, analyzed the motivation to adopt news algorithms using field theory. They identified two main "heteronomous" (p. 430) forces to the field: economic and political forces. Traditional news media lose advertisers and consumers to online platforms which puts enormous economic pressure on them. The automation trend is one of the reasons for the shifting of the audiences. So, automation is both reason for the economic problems news companies are facing, and possible solution. Obviously, Automated News can save a lot of money. The cost of adopting an algorithm is not seen as a burden by executives (Kim & Kim, 2017a). Some news editors even hope that automation helps journalism become more

independent. Automation use helps reduce costs and generate higher revenues. This combination can potentially lead to the reduction of reliance of news agencies on advertisers and governments (Wu, Tandoc, & Salmon, 2019).

2.5 Journalists' opinions on algorithms

Obviously, journalists have an opinion on Automated News as well. Journalists' attitude towards algorithms is one of the most researched topics in the context of Automated News. In general, they tend to agree that algorithms are already changing the field and will change it even more in the future. Journalists themselves describe Automated News as a phenomenon that will influence journalism heavily beyond sports news. There are two main reasons why Automated News must be taken seriously. Firstly, there is a trend to produce more pieces in less time and at a lower cost. The motivations for that were discussed in the previous chapter 2.4. Algorithms do exactly that. And secondly, the quality of articles Automated News are competing with is not that high. Right now, Algorithms are not used to compete with high quality news in large national newspapers that put a lot of emphasis on analysis, detail and background information. They mostly compete with low quality articles (often with articles from news agencies) that can be retrieved from the internet for free. For those articles news companies set the bar rather low and news algorithms can compete (Van Dalen, 2012).

Because they know that algorithms so far only compete with low-quality articles, most journalists do not appear to be concerned. They do know about the limitations of Automated News. According to Thurmann, Dörr and Kunert, 2017, journalists mostly identify the following limitations: They rely on single data streams, that are mostly one-dimensional. They can't question the data. There is no human angle in the texts. They can't work creatively. Some of the limitations, that are often mentioned, can be overcome and will be overcome in the future as the algorithms get better. One main criticism is that algorithms usually don't write about the historic context. That can and will be overcome in the future.

However, it is doubtful that algorithms will ever be able to analyze and interpret data the same way humans can (ibid.).

In a different study, journalists were asked about their strengths and weaknesses in an environment where News Algorithms are present. The strengths mentioned are: Creativity to go beyond clichés and add humor. Flexibility to write more in-depth pieces. Analytical skills that allow humans to go beyond description of data. Journalists' weaknesses in this context are: Higher costs, inability to provide the same breadth of coverage and lower speed when producing content (Van-Dalen, 2012). This has two effects: On the one hand, automation can help free up journalists and take away the burden of doing routine tasks like monitoring stories, writing low-level stories and uploading stories. Instead they get time to work on deeper stories and get more time for research. On the other hand, journalists who do only routine tasks are at risk of losing their jobs. Additionally, automation removes the need to hire journalists for low-level reporting (Van-Dalen, 2012; Wu, Tandoc, & Salmon, 2019). On the one hand that is especially problematic for young journalists, trying to find a job in journalism. Beginners are mostly asked to do routine tasks to start their careers to gain some experience (Linden, 2017b). On the other hand, optimistic journalists also see it as an opportunity that Automated News pushes them to do a better job. It is also seen as positive that stories that go uncovered now because they are only local stories can be covered with the help of algorithms (Van-Dalen, 2012).

Kim & Kim, 2017a, identified three groups of people among journalists. Type 1 has minor concerns. This group believes that Automated News have too many limitations to be a real threat for them. They do not think that newspapers' earnings will increase dramatically after the adoption of Automated News. Type 2 believes that Automated News are a threat for the social status of journalists. They do not trust their executives and express skepticism towards their promises of job security. Type 3 takes a positive stance towards the introduction of all

new forms of journalism. They believe that Automated News will help create news stories in entirely new ways. According to Wu, Tandoc, & Salmon, 2019, younger journalists have less problems using new technologies like news algorithms, whereas older journalists with a lot of experience feel threatened that technology will take over portions of the jobs they used to do.

This notion is backed by the interviews that Wu, Tandoc & Salmon, 2019, did. The hiring requirements of newsrooms are ever changing. In the future, news agencies could prioritize coding or data analysis as a value-added, or even essential, skill. Within the last ten years or so, photography and video skills became essential assets for journalists. Coding and using data might be the next big assets for journalists in the next few years.

2.6 Ethical challenges

Obviously, the use of Automated News comes with a lot of ethical challenges. Dörr and Hollnbuchner, 2017, distinguish between challenges during the input, throughput and output of the news. Most challenges arise in the input stage when the data is collected and "handed" to the algorithms. The data must be accurate and reliable for it to be used in journalistic articles. However, that is hard to guarantee in a lot of cases. The algorithm can only work with the data in the database. So, if the data is incomplete, it will change the result of the algorithm significantly. Other important questions are: Who collected the data? How was the data collected? Does the data meet the prerequisite of objectivity? Readers must know about that so they can decide for themselves if they want to read the story and if they want to rely on the information that is given to them. Another important question is whether personal rights are infringed during the collection.

In a different study some more factors were added. It is also important for the audience to know when the data was collected, how large the dataset is and whether it was somehow edited and if so, how it was edited. Audiences also might want to know about the collection

method and its limitations. However, this information is not disclosed in most cases. There are two main reasons for that. Firstly, there is no real incentive from a business standpoint in disclosing that information. Secondly, news companies are concerned they might overwhelm the audience with too much information (Diakopoulos & Koliska, 2017).

Another interesting question is whether the algorithmic nature of an article should be disclosed in bylines and what should happen if a piece of Automated News was used by a human writer. Research suggests that there is a huge gap between what journalists think should be disclosed and what gets disclosed. Montal and Reich, 2017, came up with a list of things that should be done: The byline should be attributed to the programmer if he or she is an in-house programmer. The algorithmic nature of the piece should be disclosed, and the used data sets should be described. If collaborative work is done between humans and algorithms, the byline should be attributed to humans. The full disclosure should then declare the objects in the piece that were created by an algorithm.

Lewis, Sanders, & Carmody, 2019, thought even further than just transparency issues and asked the interesting question, whether Automated News should be granted First Amendment protection in the United States. Obviously, this question can be extended to the interpretation of press freedom in all legal systems around the world. Right now, there is no consensus whether Automated News should enjoy that type of protection. In general, the US Supreme court has extended protection to content produced using emerging technologies. It usually takes until the technologies have been adopted within society, which might not be the case with Automated News just yet. Usually in cases of accusation of libel commited by journalists, the plaintiffs need to proof that the defendants were aware that they were publishing something false or at least showed reckless disregard of the truth. In the US legal system, it is usually very hard to proof that. Lewis, Sanders & Carmody, 2019, believe that plaintiffs would need to proof “that the programmer knew, or should have known, that the

algorithm would produce false statements" (p. 68). Alternatively, plaintiffs could provide evidence, "that the editor or publisher knew, or should have known, that the algorithm was not foolproof" (p. 69). To sum up, the legal implications of Automated News are still not entirely clear, just like the ethical issues are not met with the transparency scholars call for.

2.7 Consumers perception of Automated News

The main objective of this paper is to research the consumers perception of Automated News. Since Automated News are a rather new phenomenon there is only little research on that topic. To the authors knowledge there are only very few articles about the perception of Automated News. Those papers mostly agree that Automated News score higher in credibility but not as high in the category well-written. It must be mentioned that it is very hard for consumers to correctly identify whether a piece was written by a human or by an algorithm. The differences in the perception are usually very low, often not significant.

2.7.1 Credibility of Automated News

A lot of the studies were done using expectation theory as their theoretical framework. Consumers have expectations when they read articles. Haim and Graefe, 2017, found out that the expectations for human-written news are too high and expectations for computer-generated news are too low. People then get disappointed by human-written news and they rated them lower in some categories in Haim and Graefe's research. Interestingly, people already expected Automated News to be more credible. This could be because people think that algorithms are able to tell the facts without any bias. This was confirmed by Liu & Wei, 2018, who found that computer-written news were perceived as significantly more objective. The two authors investigated how trust in the news organization that uses computer-written news influences the perception. According to their findings, computer-written news are always considered more objective than human-written news regardless of the news organization. However, the trust in the medium does play a role. If a news medium is trusted

less, the computer-written news published by that medium is also trusted less (ibid.). Clerwell, 2014, went a little more in-depth and found out that Automated News are considered more descriptive, informative, trustworthy and objective. He then concluded 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?” (Clerwell, 2014, 526)

As already mentioned, computer-written news often get edited by humans before publication. There are only few studies on tandem authorship between computers and humans. One of the few studies was done by Waddell, 2019. He tested which effects the three different possibilities (human authorship, machine authorship, tandem authorship) cause. Both machine and tandem authors were considered less biased than human authors. Not suprisingly, machine authors were perceived less anthropomorphic. Both bias and anthropomorphism played a role in the perception of credibility. Credibility is increased by the notion that computers are less biased but is decreased through the pathway of anthropomorphism. This could be an explanation why the perceived differences are usually very low (ibid.). Waddell also found similar results in a different paper. People expect algorithms to be less biased. Those high expections get undermined and computer-written news get evaluated less positively. It is important to note that participants who were more familiar with robots from popular media were less likely to rate computer-written articles negatively in terms of its credibility. It is important to note that Waddell only focused on source attribution and not on the content of the article itself (Waddell, 2018).

2.7.2 Readability of human-written news

Human written articles on the other hand are considered livelier and more readable. Differences in all categories were rather low however, which suggests that computer-generated news can compete with human-written news. Once again, source expectations seem to play a role. Graefe, Haim, Haarmann, & Brosius, 2016, let people read different news

articles on sports and finance and told them a source of the article. However, they altered the sources a lot so some people, who were reading human-written stories, thought they were reading computer-written news and vice versa. People enjoyed reading stories that were declared as human-written a lot more. However, the participants in their study perceived Automated News (both declared as computer-written and human-written) as more credible (ibid.). Clerwell, 2014, found out that human-written news items are considered more coherent, well-written, clear and pleasant to read.

Waddell found out that that human authors are perceived as more anthropomorphic. Tandem authors are also seen as more anthropomorphic than machine authors. He concluded that "news writing is still largely perceived as ‚a human's job'". (Waddell, 2018, p. 248) He suggested that readers still appreciate news written by humans more and only react positively to news selected by algorithms (Waddell, 2018; Waddell, 2019). A possible explanation is offered by Liu & Wei, 2018. In their study, participants regarded human writers as having significantly more expertise.

2.7.3 Other factors that influence the different perceptions

It has to be mentioned again that the difference in perceptions is usually low and often not significant. Van der Kaa & Krahmer, 2014, found no significant differences between the percpetion of computer-written and human-written news. However, the general rule that computer-written news are considered more credible whereas human-written news are considered more readable is established in the literature.

2.7.3.1 Expectation theory

As described above, expectations play a role. But expectations are not the only explanation for the differences. In some studies (e.g. Clerwell, 2014) the source of the computer-written and human-written articles was not disclosed to participants. In the study done by Graefe,

Haim, Haarmann, & Brosius, 2016, computer-written articles were rated as more credible than human-written articles even if computer-written articles were wrongly primed as human-written and vice versa. They also found out that human-written articles, wrongly primed as computer-written, were pereceived as more readable than computer-written articles primed as human-written. So, while expectation theory is often used to explain the different perceptions, there have to be additional factors that play a role.

2.7.3.2 Culture

Zheng, Zhong, & Yang, 2018, performed a cross-cultural study in the United States and China. Their results indicate that culture did not have an influence on the different perception. While participants from the United States perceived the quality of the news articles the authors presented to them as higher in general, there was no difference within the US sample or within the chinese sample. It also did not matter if the news were primed as written by the *New York Times*, a well-known traditional newspaper, or by *Buzzfeed News*, a rather new online medium (ibid.). This again shows that expectation theory can't be used to explain the differences in perception without taking the content itself into account.

2.7.3.3 Type of news article

One factor that could make a difference, is the type of news article that is researched. In one study, spot news written by machines were perceived as significantly less credible than interpretive news. Interestingly, interpretive news tended to be more credible when written by computers rather than humans (Liu & Wei, 2018). That is a remarkable result, given that the most popular approach right now is to use algorithms to produce short, low quality pieces, whereas humans write longer, interpretive pieces (Van-Dalen, 2012; Wu, Tandoc, & Salmon, 2019; Linden, 2017b).

2.7.3.4 Topic

One study offered another possible factor that could have an influence. Wölker & Powell, 2018, found out that the topic of the news could make a difference. They investigated how credible human-written and computer-generated news are. In their online survey they found out that automated content scored significantly higher than human content for sports articles. However, no significant difference was detected for finance articles. Combined pieces, that were written by humans with the help of algorithms, are considered credible. Haim & Graefe, 2017, also tested if the topic makes a difference but they found no differences across different topics.

2.7.4 Outlook

The idea, that the topic can have an influence on the different perception of computer-written and human-written news has only been researched rarely. However, the idea that the topic can have an effect is worth investing and will be investigated in this paper. As pointed out by a lot of scholars, Automated News is here to stay. It already has an impact on the field of journalism, one of the most important institutions in modern democracies, and it will have an even bigger impact in the future. Hence, Automated News should be researched as thoroughly as possible so that the scientific community knows as much as possible about Automated News to educate (future) journalists, producers and society as a whole about that topic.

Obviously, the success of Automated News will largely depend on the perception of the consumers. Kim & Kim, 2017, wrote that, right now, the willingness by consumers to accept computer-generated news is more important to producers than saving costs. So even producers seem to agree, that the success of Automated News will mainly depend on the consumers perception. So, doing research on that topic is not only important for the scientific community and society but also for producers.

Finally, this topic is also important for developers of news algorithms. There are a few companies around the world that specialize in producing this kind of software. One German-speaking company, Retresco, even won a United Nations World Summits Award[1] in the category Business & Commerce for their news algorithm.

This all shows that knowledge about consumer's perception is relevant in today's research about Journalism and Innovations in the Media. This paper attempts to contribute to a better understanding of influences on the perception of computer-written news.

2.8 Cognitive Elaboration

One more interesting question in this regard, that has only been researched once to the author's knowledge, is how human- or machine-authorship influence the cognitive elaboration of readers. According to Gil de Zúñiga, 2017, „the concept of the elaboration […] refers to the process through which individuals […] think and reflect on that which was discussed and relate it to their environment and their own experiences." (p. 66) A definition is provided by Eveland Jr., 2001: „Elaboration is the process of connecting new information to other information stored in memory, including prior knowledge, personal experiences, or the connection of two new bits of information together in new ways." (p. 573) To sum up, cognitive elaboration is an indicator of how many mental effort people spend, whenever they process new information (Zhang, Fang, Wei, & He, 2013).

When it comes to human-written and computer-written news, computer-written news led to less cognitive elaboration compared to human-written news in the study by Liu & Wei, 2018. This is the only finding the author was able to find on the different cognitive elaboration caused by human-written and computer-written news. Obviously, the news industry's main

[1] https://www.worldsummitawards.org/winner/retresco/ (2019-07-09)

goal is not always achieving a high cognitive elaboration. But it is still a good indicator how engaged consumers are with the content (ibid.).

3 Study

3.1 Research Questions and Hypotheses

In this paper, the influence of the topic on the different credibility, readability and cognitive elaboration between computer-written and human-written news pieces is researched. This chapter is therefor divided into the three subchapters credibility, readability and cognitive elaboration, which will serve as the dependent variables in this study. The independent variables are human- versus computer-authorship.

3.1.1 Credibility of computer-written and human-written news

Research Question 1: What difference does the topic make when comparing human written and computer-generated news items in terms of their credibility?

As mentioned above Wölker and Powell, 2018, found out that the difference in the perception of computer-written and human-written news could depend on the topic. However, that was not their main discovery. It was rather a side note, albeit a very interesting one that is worth researching. Finding out if this initial finding can be confirmed in a study dedicated to the idea of different differences between computer written and human written news is the main goal of this piece. Following Wölker's and Powell's findings the corresponding hypothesis are:

Hypothesis 1a: Computer generated news about sports are rated significantly more credible than human written news.

Hypothesis 1b: There is no significant difference between computer generated and human written news on financial topics in terms of their credibility.

Hypothesis 1c: Computer generated weather reports are considered more credible than human written weather reports.

Research Question 1 can only be answered when all hypotheses 1a, 1b and 1c are considered. It seems worth clarifying that Research Question 1 does not ask for a difference in the perception between computer written and human written news in general. This coarse-grained question of whether there is a difference has already been discussed in previous research. Research Question 1 is more fine-grained and asks for a specific detail. This piece is not about asking if there is a difference between computer written and human written news. It asks for factors that influence that difference, specifically for the difference in the difference across different topics. Therefore, taking only one of the three hypotheses into account would not be enough. We need to consider all hypotheses to find out about the differences between computer written and human written news across different topics and to answer the research question.

Wölker and Powell only compared sports and finance news. So, we do not know anything about differences in the different perception of other topics. However, the topic of weather reports will also be covered in this piece to research the field a little more detailed. And since there is no research on those differences to this point, the hypotheses follow the general trend that has been established in the literature: computer written news are more credible, human written news are more readable.

3.1.2 Readabilty of computer-written and human-written news

Research Question 2: What difference does the topic make when comparing human written and computer-generated news items in terms of their readability?

Since Wölker and Powell, 2018, only focused on differences in the perception of credibility and since Van der Kaa and Kraehmer 2014 and Graefe et al 2016 did not distinguish between the topic, it is hard to predict whether or not there are differences in the perception of readability when comparing the different perception between computer written and human written news. Since nobody has noticed such differences to this point, we must assume that computer written and human written news in those two fields are perceived equally different in the perception of readability. Hypothesis 2a, 2b and 2c are therefore:

Hypothesis 2a: Human written sports news items are considered significantly more readable than computer written sports news.

Hypothesis 2b: Human written finance news items are considered significantly more readable than computer written finance news.

Hypothesis 2c: Human written weather reports are considered more readable than computer written weather reports.

3.1.3 Influence of the authorship on cognitive elaboration

Research Question 3: What influence does the authorship have on the cognitive elaboration?

This paper is not just about the differences in perception but also about the difference in cognitive elaboration. As described in chapter 2.8, the authorship of humans versus computers can make an influence on how the article is perceived. Following the initial findings of Liu & Wei, 2018, the corresponding hypothesis are:

Hypothesis 3a: Cognitive elaboration is higher when participants read human-written sports news compared to computer-written sports news.

Hypothesis 3b: Cognitive elaboration is higher when participants read human-written finance news compared to computer-written finance news.

Hypothesis 3c: Cognitive elaboration is higher when participants read human-written weather news compared to computer-written weather news.

3.2 Variables

There are some words that need to be defined explicitly. The topics sports news, finance news and weather reports seem pretty self-explanatory. Human written and Computer written news also do not need a lot of clarification. A human written news article is an article that is solely written by a human. A computer written article is an article that was produced by a computer automatically.

It is a little bit tougher to define credible and readable. In previous studies about the different perception of computer written and human written articles those two words were defined

using a few different characteristics. Graefe, Haim, Haarmann, & Brosius, 2016, followed Sundar, 1999, who defined 17 items to measure the quality of news. Graefe et al then added four more items that were used by Van der Kaa and Krahmar (2014) and by Clerwall (2014). Of those 21 words in total, Graefe et al used twelve to describe the concepts of credibility, readability and journalistic expertise, with four words representing each of the concepts. Since journalistic expertise is not relevant in this study, only credibility and readability must be defined. The author decided to follow Graefe et al to make the results comparable. Credibility is defined as the perception of the participants on how accurate, trustworthy, fair and reliable the news items are. Readability is defined as the perception on how entertaining, interesting, vivid and well-written the articles are.

3.3 Method

To test the hypotheses and answer the research questions, the author conducted an experiment. The experiment, which was spread as an online survey using the website www.soscisurvey.de, started with some questions about the media usage of the participants to break the ice. In the main part of the experiment participants were asked to read three different texts and to rate them.

3.3.1 Stimulus Material

In total, there were six articles in this experiment. To test the differences between human-written and computer-written news articles on sports, finance and weather, there was one article representing each one of the categories. The author tried to choose rather similar articles. To clearly tell whether an article was written by a human or a computer the author looked for articles where human or computer authorship is clearly stated in the byline. When it comes to sports articles it was tougher to find material that was clearly written by a computer than it was to find human-written articles. After some research, the author found out that the *Washington Post* now uses computer-written articles to preview Baseball, Ice-hockey

and College Basketball games. They use articles generated using technology from *Hero Sports* and data from *Sportradar*. Since the author of this paper is based in Vienna, he decided to go with an Ice-Hockey article. Of the three sports mentioned above, hockey is by far the most popular in Vienna. The author picked the preview of a game between the San Jose Sharks and the St. Louis Blues on May 16th, 2019. As a human written sports article, the author picked the preview of the same game written by Ryan Bristlon of wsn.com.

When it comes to finance articles, the main challenge was to find similar articles written by algorithms and humans. All earning's reports that are published by the *Associated Press* are now generated by algorithms. It was very easy to find some of them. It was also easy to find analytical human-written pieces. It is worth noting at this point that the idea, that was already described in this paper, that algorithms are used to write short reports about the hard facts whereas humans are tasked with writing in-depth analytical pieces seems to be well-established when it comes to financial news. However, the author was finally able to find a human-written earnings report of *EBay* written by Claudia Assis for marketwatch.com. This report was used as an example for a human-written financial article. As a computer-written article, the author chose an *EBay* earnings report generated by an algorithm for the *Associated press*.

The toughest part was finding a human-written weather report with a byline. The author finally found one in the British Tabloid *The Sun* written by Jenny Awford. As a computer-written example, the author used a weather report for the UK generated by Retresco, the German company that was already mentioned before in chapter 2.7.4. It has to be mentioned that the two weather reports appear rather different to each other to the author, but the author was not able to find more similar examples.

All six articles can be found in full length in the appendix.

3.3.2 Experiment

In the experiment, each participant was only asked to rate three of the six texts to prevent the experiment from taking too long. It was decided completely randomly which of the six articles the participants were shown and the order of the three chosen articles was also randomized to prevent effects caused by the order. Since Graefe at al, 2016, found out that declaring articles wrongly as human written or computer written makes no significant difference, the source of the article was not disclosed at all. This piece is only about the perception of the content and not about expectations. So, the source of the article is not disclosed to avoid an influence on the participants.

After reading each of the three articles, that were chosen for the participants, they were asked to rate the articles in terms of credibility and readability. As it was mentioned above this was done the same way as in Graefe's et al, 2016, study. They defined credibility as accurate, trustworthy, fair and reliable. And they defined readable as entertaining, interesting, vivid and well-written. So after reading each article the participants were asked to rate it on a seven point scale, which is the same scale as Graefe at al used, based on the eight characteristics accurate, trustworthy, fair, reliable, entertaining, interesting, vivid and well-written, with four of those words representing the concept of credibility and four representing readability.

Since Cognitive elaboration is about measuring how much engaged participants are when reading the articles, and how much they try to connect the information with their past experiences, they were asked to rate how much they agreed the following four statements: "I thought about the content of this article"; "I thought about similar news articles I read in the past"; "I tried to connect the content of this article to my knowledge about the topic."; "The topic of this article is interesting to me".

For each of the three topics we now have three different groups we can consider. There are some participants that were only asked to rate the computer-written article, some that were only shown the human-written article and some that were shown both articles on the topic. This will be considered in chapter 4 where the results are described.

3.3.3 Acquisition of Participants

Participants in the experiment were acquired through different ways to make sure the sample size was large enough. The author started by sending the link to the online survey to his fellow students and friends and asked them to share the link with three people the author did not know. The author also went around campus at the University of Vienna and asked people in person to fill out the survey. In order to do that the author printed out the link and a QR code to make sure that interested people were able to open the survey quickly. Once they opened the link the author left to make sure that the participants were not influenced by his presence. The third way to acquire participants was using the platform SurveyCircle[2] and its Facebook page. The concept there is very easy: People looking for participants in their studies share the links there and participate in each other's studies. The author participated in over 100 other studies, hoping that the authors of those studies would participate in his experiment. Obviously, it is hard to say how many participants were acquired through each way as the study was anonymous. The author was only able to keep track of the total number of participants. In total, 524 people opened the link to the survey. By looking at the timely development of that number, it can be estimated that about 150 to 200 participants were acquired through the snowball system. Another 80-100 were acquired on SurveyCircle and its Facebook page. The rest was acquired in person.

[2] www.surveycircle.com

3.3.4 Sample

In total, 524 people opened the link to the survey and answered at least the first question. 415 people (79.2 %) completed the survey. Of the 415 participants, 135 (32.5 %) reported being male, 272 (65.5 %) reported being female, 2 reported having a non-binary gender and 6 did not want to disclose their gender. The mean age of the participants was 24.4 years (SD: 5.12).

211 (50.8 %) participants reported living in Vienna, 70 (16.9 %) in Lower Austria (Area around Vienna), 12 (8.3 %) in other regions of Austria, 82 (19.8 %) in Germany and 40 (9.6 %) in countries other than Austria and Germany. 194 (46.7 %) of the sample had at least a Bachelor's degree. The majority of the sample, 248 participants (59.8 %), were students, 142 (34.2 %) of the entire sample reported that they were either working full-time (70) or part-time (72). Since the experiment was held in English it is worth mentioning that 28 (6.7 %) reported that their native language was English. 336 (81.0 %) of the participants were native speakers of German. Other native languages included Bosnian/Serbian/Croatian (26 participants, 6.3 %). Hungarian, Russian, French and Spanish were the native language of between five and ten participants. The number of all other native languages reported was under five.

4 Results

As explained above in chapter 3.3.2, all participants were asked to rate three out of the six articles which were chosen completely randomly. This means that some participants were asked to rate both the machine-written and human-written articles on one topic. Others were asked to rate only one of the two articles on one topic. Hence, there are two groups that need to be considered. The group that read only one article on a topic will be referred to as "Group A" from now on. The group that read both articles will be called "Group B". As the main topic of this paper is to investigate the influence of the different topics, this chapter will be sorted after the three topics sports, business and weather and in each of those subchapters the results are presented separately.

To compare the tendencies between the groups that only rated one of the two articles on a topic on a 7-point Lickert scale a Mann Whitney U Test was performed. To compare the tendencies of the group that rated both articles on a topic a Wilcoxon Test was performed. On the following pages only the most important attributes will be presented in tables for the sake of clarity. More detailed tables can be found in the appendix.

4.1 Sports Articles

As illustrated in Table 1, Hypothesis 1a was partially supported in Group A, the group of people that rated only one of the two articles. The computer-written article was rated significantly more accurate ($p = 0.005$, $r = 0.18$) and reliable ($p = 0.028$, $r = 0.14$). However, the effects were rather small as indicated by the values of r. No significant difference was found for the attributes trustworthy ($p = 0.205$) and fair ($p = 0.120$). Partial support was also found for Hypothesis 2a. The human-written sports article was rated significantly more vivid ($p = 0.001$, $r = 0.20$) and scored higher on the attribute well-written ($p = 0.004$, $r = 0.18$). No

significant difference was found for the attributes entertaining ($p = 0.251$) and interesting ($p = 0.576$).

Hypothesis 3a was not supported at all. Human authorship did not lead to higher cognitive elaboration. Three of the four statements to measure cognitive elaboration were rated very similarly for computer written and human written news. The only statement that was rated differently was “I thought about similar news articles I read in the past”. Interestingly, in contrast to hypothesis 3a, which predicted that human authorship would lead to higher cognitive elaboration, this statement was rated higher when participants read a computer-written article. The significance here is very low though ($p = 0.074$, $r = 0.11$).

Table 1

Comparison of the credibility attributes between participants who rated only one sports article

Attribute	Computer-written		Human-written		Mann-Whitney U	p	r
	N	Mean Rank	N	Mean rank			
Accurate	127	137.07	122	112.41	6213.50	0.005	0.18*
Trustworthy	127	131.01	123	119.97	7108.00	0.205	-
Fair	126	131.19	122	117.59	6842.50	0.120	-
Reliable	124	133.14	122	113.70	6369.00	0.028	0.14*
Entertaining	127	120.41	123	130.75	7164.50	0.251	-
Interesting	127	123.03	123	128.05	7496.50	0.576	-
Vivid	126	110.15	122	138.66	5861.00	0.001	0.20*
Well-written	124	112.94	122	138.46	6216.00	0.004	0.18*
I thought about the content of this article.	127	125.57	123	125.42	7802.00	0.988	-
I thought about similar news articles I read in the past.	127	133.32	123	117.43	6817.50	0.074	0.11*
I tried to connect the content of this article to my knowledge about the topic.	127	122.09	122	128.03	7377.00	0.509	-
The content of this article is interesting to me.	127	124.26	123	126.78	7653.00	0.776	-

Note: * Small effect (0.1 < r < 0.3). ** Medium effect (0.3 < r < 0.5). *** Large effect (r > 0.5).

As illustrated in table 2, the effects were a little bit different in Group B, the group that got to read both articles. Hypothesis 1a was not supported. Like in the other group, there was no different perception of the article's trustworthiness ($p = 0.528$). In contrast to the other group, there was no significant difference measured regarding the article's accuracy ($p = 0.353$). Interestingly, unlike in the other group, there was a significant difference with a medium effect regarding the article's fairness ($p = 0.009$, $r = 0.30$). That difference supports Hypothesis 1a. However, in vast contrast to Hypothesis 1a, the reliability of the human-written article was rated higher. The significance of that difference is low ($p = 0.058$, $r = 0.21$).

Table 2

Comparison of the credibility attributes between participants who rated both sports articles

Attribute	Computer-written		Human-written		Total	p	r
	N	Mean Rank	N	Mean rank			
Accurate	25	20.04	16	22.50	79	0.353	-
Trustworthy	28	21.32	18	26.89	78	0.528	-
Fair	30	22.92	13	19.98	78	0.009	0.30**
Reliable	31	21.97	14	25.29	77	0.058	0.21*
Entertaining	23	22.96	28	28.50	77	0.200	-
Interesting	21	21.48	26	26.04	78	0.224	-
Vivid	17	20.47	35	29.43	76	0.002	0.35**
Well-written	19	25.61	35	28.53	78	0.025	0.26**
I thought about the content of this article.	17	27.26	32	23.80	77	0.126	-
I thought about similar news articles I read in the past.	21	22.69	19	18.08	78	0.365	-
I tried to connect the content of this article to my knowledge about the topic.	22	24.09	27	25.75	76	0.399	-
The content of this article is interesting to me.	15	16.10	18	17.75	78	0.475	-

Note: * Small effect (0.1 < r < 0.25). ** Medium effect (0.25 < r < 0.4). *** Large effect (r > 0.4).

Hypothesis 2a was partially supported. The articles were not rated significantly different in terms of the attributes entertaining ($p = 0.200$) and interesting ($p = 0.224$). The different ratings between the attributes vivid ($p = 0.002$, $r = 0.35$) and well-written ($p = 0.025$, $r = 0.26$) were also found in Group B. The r values were higher in this group indicating that the effects are medium rather than small. Hypothesis 3a was also not supported in this group. There was no significant difference between the cognitive elaboration when reading the two articles.

4.2 Finance news

As illustrated in table 3, Hypothesis 1b, which predicts that there would be no significant differences in terms of the articles' credibility, was not supported in Group A of the two finance articles. For three of the four attributes describing credibility a significant difference was detected. Interestingly, the general rule that computer-written articles are considered more credible was not measured in this experiment. The human written article was rated significantly higher in terms of accuracy ($p = 0.011$, $r = 0.16$), trustworthiness ($p = 0.033$, $r = 0.13$) and fairness ($p = 0.002$, $r = 0.19$). The r values indicate that the effects are significant but rather small. No significant difference was measured between the articles' reliability ($p = 0.127$).

Hypothesis 2b was partially supported. There was a significant difference reported by the participants in terms of the attribute well-written ($p = 0.021$, $r = 0.15$). The human-written article was also perceived to be more vivid ($p = 0.066$, $r = 0.12$), although the significance is very low. No significant difference was measured between the attributes entertaining ($p = 0.283$) and interesting ($p = 0.498$).

Hypothesis 3b was not supported at all. No significant difference was found for either of the four statements measuring cognitive elaboration.

Table 3

Comparison of the credibility attributes between participants who rated only one finance article

Attribute	Computer-written		Human-written		Mann-Whitney U	p	r
	N	Mean Rank	N	Mean rank			
Accurate	122	114.39	129	136.98	6452.00	0.011	0.16*
Trustworthy	123	116.27	128	135.35	6675.50	0.033	0.13*
Fair	123	112.37	128	139.09	6196.00	0.002	0.19*
Reliable	123	120.00	130	133.62	7134.00	0.127	-
Entertaining	123	122.04	130	131.69	7385.00	0.283	-
Interesting	123	123.36	129	129.49	7547.50	0.498	-
Vivid	121	115.22	125	131.51	6561.00	0.066	0.12*
Well-written	123	115.83	129	136.67	6621.00	0.021	0.15*
I thought about the content of this article.	123	122.54	128	129.32	7446.50	0.453	-
I thought about similar news articles I read in the past.	123	122.36	129	130.45	7424.50	0.371	-
I tried to connect the content of this article to my knowledge about the topic.	123	123.74	129	129.13	7594.00	0.551	-
The content of this article is interesting to me.	123	122.83	128	129.05	7481.50	0.490	-

Note: * Small effect (0.1 < r < 0.3). ** Medium effect (0.3 < r < 0.5). *** Large effect (r > 0.5).

In Group B, no significant differences were measured. Hypothesis 1b was supported. There was no difference in the participant's ratings on how accurate ($p = 0.305$), trustworthy ($p = 0.532$), fair ($p = 0.724$) and reliable ($p = 0.723$) the articles were. Hypothesis 2b, predicting that the human-written article would score higher in terms of its readability, was not supported in this group. No significant differences were found for the attributes entertaining ($p = 0.708$), interesting ($p = 0.104$), vivid ($p = 0.806$) and well-written ($p = 0.199$). Hypothesis 3b was not supported either. There was no significant difference measured between the ratings of the participants on the four statements measuring cognitive elaboration.

Table 4

Comparison of the credibility attributes between participants who rated both finance articles

Attribute	Negative Ranks		Positive Ranks		Total	p	r
	N	Mean Rank	N	Mean rank			
Accurate	15	17.03	20	18.73	68	0.305	-
Trustworthy	17	18.32	20	19.58	69	0.532	-
Fair	18	17.61	16	17.38	68	0.724	-
Reliable	18	17.64	16	17.34	69	0.723	-
Entertaining	23	21.09	22	25.00	68	0.708	-
Interesting	15	18.47	24	20.96	68	0.104	-
Vivid	20	22.65	23	21.43	69	0.806	-
Well-written	24	18.50	23	29.74	69	0.199	-
I thought about the content of this article.	16	15.53	16	17.47	69	0.766	-
I thought about similar news articles I read in the past.	20	18.63	17	19.44	69	0.747	-
I tried to connect the content of this article to my knowledge about the topic.	20	19.02	18	20.03	69	0.882	-
The content of this article is interesting to me.	15	16.40	21	20.00	69	0.154	-

Note: * Small effect ($0.1 < r < 0.25$). ** Medium effect ($0.25 < r < 0.4$). *** Large effect ($r > 0.4$).

4.3 Weather reports

As illustrated in table 5, the were some highly significant differences measured in the ratings of the two weather reports between Group A. Hypothesis 1c was mostly supported. The computer-written article was rated highly significantly more trustworthy ($p = 0.000$, $r = 0.59$) and more reliable ($p = 0.000$, $r = 0.70$). The effect size in those two cases is large as evidenced by the r values. The computer-written article also scored higher in terms of fairness ($p = 0.090$, $r = 0.11$). The difference is significant, but the effect is only small. The one big outlier here is the rating of the accuracy of the article. In complete contrast to hypothesis 1c, the human-written article was rated significantly more accurate ($p = 0.000$, $r = 0.29$). The effect size of this difference is still small, but on the border to being considered medium.

Hypothesis 2c was supported in this study. The participants thought that the human-written weather report was significantly more entertaining ($p = 0.000$, $r = 0.85$) with a large effect size. The human-written article was also considered significantly more interesting ($p = 0.000$, $r = 0.31$) and well-written ($p = 0.000$, $r = 0.33$), both having a medium effect. The human-written article was also rated higher in terms of its vividness, but not significantly ($p = 0.109$). Hypothesis 3c was not supported. There were some effects and some significant differences but there was no clear direction. Participants rated the statement “I thought about the content of this article” significantly higher after reading the human-written article ($p = 0.000$, $r = 0.52$). As indicated by the rather large value of r, the effect is large. The statement “The content of this article is interesting to me” was also rated significantly higher after reading the human-written article ($p = 0.002$, $r = 0.20$). The effect size was small. In contrast to those two findings, the participants significantly agreed more with the statement “I tried to connect the content of this article to my knowledge about the topic” after reading the computer-written

article ($p = 0.000$, $r = 0.38$). No significant difference was found in the participants' rating of the statement "I thought about similar news articles I read in the past" ($p = 0.401$).

Table 5

Comparison of the credibility attributes between participants who rated only one weather article

Attribute	Computer-written		Human-written		Mann-Whitney U	p	r
	N	Mean Rank	N	Mean rank			
Accurate	136	109.50	114	144.59	5576.00	0.000	0.29*
Trustworthy	136	159.50	114	84.94	3128.00	0.000	0.59***
Fair	136	131.50	114	118.34	6936.00	0.090	0.11*
Reliable	136	165.00	114	78.38	2380.00	0.000	0.70***
Entertaining	136	75.50	114	185.15	952.00	0.000	0.85***
Interesting	136	107.50	114	146.97	5304.00	0.000	0.31**
Vivid	136	119.50	113	131.62	6963.00	0.109	-
Well-written	136	144.50	114	102.83	5168.00	0.000	0.33**
I thought about the content of this article.	136	94.50	113	161.71	3536.00	0.000	0.52***
I thought about similar news articles I read in the past.	136	128.00	113	121.39	7276.00	0.401	-
I tried to connect the content of this article to my knowledge about the topic.	136	147.00	113	98.52	4692.00	0.000	0.38**
The content of this article is interesting to me.	136	114.00	113	138.24	6188.00	0.002	0.20*

Note: * Small effect (0.1 < r < 0.3). ** Medium effect (0.3 < r < 0.5). *** Large effect (r > 0.5).

In Group B, Hypothesis 1c was mostly supported. The participants considered the computer-written article significantly more accurate ($p = 0.035$, $r = 0.22$), more trustworthy ($p = 0.008$, $r = 0.28$) and more reliable ($p = 0.010$, $r = 0.27$). The effect sizes in those three cases were all small to medium. No significant difference was measured in the participants' ratings of the articles' fairness ($p = 0.707$).

Hypothesis 2c was not supported. The only significant difference that was measured was in the articles' vividness. Participants considered the human-written article significantly more vivid ($p = 0.007$, $r = 0.28$) with medium effect size. The participants rated the human-written article more entertaining ($p = 0.138$), but the computer-written article more interesting ($p = 0.109$) and well-written ($p = 0.166$). Neither of those differences was statistically significant.

Hypothesis 3c was not supported in this group. The only significant difference contradicts the hypothesis. The participants agreed with the statement "I thought about similar news articles I read in the past" significantly more after reading the computer-written article ($p = 0.049$, $r = 0.21$). The effect size is small. All other three statements were not rated significantly different.

Table 6

Comparison of the credibility attributes between participants who rated both weather articles

	Negative Ranks		Positive Ranks		Total	p	r
	N	Mean Rank	N	Mean rank			
Accurate	37	32.30	23	27.61	92	0.035	0.22*
Trustworthy	41	29.06	17	30.56	92	0.008	0.28**
Fair	34	27.46	25	33.46	92	0.707	-
Reliable	43	30.98	19	32.68	92	0.010	0.27**
Entertaining	29	37.36	44	36.76	92	0.138	-
Interesting	41	33.93	26	34.12	92	0.109	-
Vivid	26	28.31	42	38.33	90	0.007	0.28**
Well-written	41	37.95	31	34.58	92	0.166	-
I thought about the content of this article.	32	28.14	29	34.16	92	0.736	-
I thought about similar news articles I read in the past.	33	29.32	21	24.64	92	0.049	0.21*
I tried to connect the content of this article to my knowledge about the topic.	25	24.02	27	28.80	92	0.406	-
The content of this article is interesting to me.	34	30.49	24	28.10	92	0.146	-

Note: * Small effect (0.1 < r < 0.25). ** Medium effect (0.25 < r < 0.4). *** Large effect (r > 0.4).

5 Discussion

In this study, an attempt was made to investigate the influence of the topic on the different perception of computer-written and human-written news. In previous studies, the rule was established that computer-written news is seen as more credible, whereas human-written news is considered more readable. In similar fashion to Graefe, Haim, Haarmann, & Brosius, 2016, Clerwall, 2014, and Van der Kaa & Krahmer, 2014, the author conducted an experiment in which participants were asked to rate computer-written and human-written articles on sports, finance and weather. The design of the experiment led to a distinction between two groups, one group that read only either the computer-written or the human-written article on one topic, and one group that read both articles on the topic. Interestingly, the results differ a little bit within those two groups.

Looking at the tables, one can quickly notice that the effects are smaller in the group, that got to read both – the computer-written and the human-written – articles on a topic (referred to as group Group B), as opposed to the group of people, that got to read only one of the two articles (Group A). This can be explained using the findings of Clerwall, 2014, and Van der Kaa & Krahmer, 2014, who concluded that differences are rather small and by the fact that people have a hard time correctly identifying if a text was written by a human or by an algorithm.

The main question to start the paper was, if the topic of the article has an influence on how differently computer-written and human-written news articles are perceived. In this study, there is some support for that idea, which was first described by Wölker & Powell, 2018. Wölker & Powell found out that computer-written sports articles were perceived as more credible than human-written sports articles. This hypothesis was partially supported in Group

A. The participants considered the computer-written preview of a hockey game more accurate and more reliable. Group B, however, considered the human-written hockey preview significantly more reliable than the computer-written one. At least, the general rule that human-written articles are perceived as more readable, was supported in this study. The human-written hockey preview received higher ratings in the categories “vivid” and “well written” in both groups.

Wölker & Powell, 2018, had the idea that the topic could make a difference, because they were not able to find significant differences in the different perception of computer-written and human-written finance articles. In this study, some of the results go even further. Not only do the results not support the general rule that computer-written articles are considered more credible than humen-written ones. The participants in this study even rated the human-written article as more credible, with significant differences in the categories “Accurate”, “Trustworthy” and “Fair” in Group A. The results were not confirmed in Group B. No significant differences between the two articles were measured there. However, the fact, that Group A rated the human-written finance article significantly more credible is still remarkable. Not only does it break the general rule on the different perception, but it is also an important finding, as algorithms are often used for generating finance articles. The *Associated Press* for example use news algorithms for their generation of earnings reports (Linden, 2017a). Concerning readability, the general rule was partially supported in this study. Human-written finance stories were perceived more vivid and better-written in Group A, but again not in Group B.

The fact, that results differ between Groups A and B can be summed up as follows: Participants rated the news articles differently when they were only shown either the computer-written or the human-written article. In the setting where the participants got to read both articles on the same topic, the differences were a lot smaller. Obviously, one has to keep

in mind that a newspaper would not publish two similar articles on one topic. So the setting that was researched in Group A is closer to how news are read in practice.

5.1 Implications for the general rule

There are now different things the results described above can mean for the general rule. Firstly, it is possible that the general rule, that computer-written articles are considered more credible and human-written articles are more readable, actually depends on the topic. The results of this study and results by Wölker & Powell, 2018, indicate that computer-written finance news are not considered more credible than human-written ones.

The second possible interpretation is, that this rule was never really true. This interpretation is supported by the notion, that differences in this study are rather small and have been small in previous studies. Some studies, like the ones by Van der Kaa & Krahmer, 2014, or the one done in a cross-cultural setting by Zheng, Zhong, & Yang, 2018, did not find significant differences in the perception. It has to be mentioned that the rule was mostly described in studies, that focused on the consumers' expectations towards human and machine authors. Most scholars in previous studies made sure that participants in their studies recalled the author correctly. Some, like Waddell, even sorted out participants who were not able to recall correctly if the article was written by a human or an algorithm (Waddell, 2018; Waddell, 2019). So expectations towards the authors certainly played a role in finding the general rule. However, as Graefe, Haim, Haarmann, & Brosius, 2016, found out, expectations cannot explain the rule alone as computer-written articles were rated more credible than human-written news, regardless of whether they were correctly primed as computer-written or wrongly primed as human-written. To sum up, it can't be ruled out at this point that the general rule was a product of expectations of the readers towards the author but does not apply when the actual content is researched.

The third possible interpretation is, that the validity depends on the individual situation and the texts, that are chosen in studies. The stimulus material for measuring differences in perception usually gets hand-picked by the authors. Authors usually try to keep the computer-written and human-written articles similar (more on the limitations this causes in chapter 6). Clerwall, 2014, for example, even made sure that the two articles he used as stimulus material had the same length. The author of this paper tried to follow this trend and looked for similar articles on sports, finance and weather. He succeeded when looking for similar sports and similar finance articles. This could also be an explanation for the low differences. However, the author was not able to find a human-written weather report with a byline that was similar to the computer-written weather report. The two weather reports are really different from each other in the way they are written. Interested readers can see for themselves as all articles, that were used as stimulus material, can be found in the appendix.

Sure enough, the perceived differences between the two weather reports were big. The computer-written article was rated significantly more trustworthy and reliable with large effect size, whereas the human-written article was rated significantly more entertaining, interesting and well-written with medium to large effect size. Big differences were only measured in Group A. Some of the differences were also measured in Group B but the effect size was not as large, which is also a very interesting observation as it has to be expected that participants noticed that they were reading two rather different weather reports. Interestingly, Group A rated the human-written article as significantly more accurate, which contradicts the rule. The notion, that a big difference in the stimulus material led to a big difference in the dependent variables, should not come as a surprise but it is still an interesting point to make, as stimulus material in past studies was mostly similar or even identical in content.

5.2 Low differences in Cognitive Elaboration

The cognitive elaboration of participants was also measured in this study. The most significant differences were found in Group A of the weather report. Participants thought about the content of the human-written article significantly more with large effect size and thought that the human-written article was significantly more interesting to them. Interestingly, that did not lead to the participants trying to connect the content of this article to their knowledge about weather. They did that significantly more when reading the computer-written article. So, the results are contradictory, but clearly the human-written weather article caused an effect on the participants cognitive elaboration. In all other different scenarios, the participants did not show different cognitive elaboration.

5.3 Implications of the findings for the field of journalism

There are some important implications for producers, editors and journalists. Firstly, they have to admit, that news algorithms fulfill their task. As Van Dalen, 2012, noted, algorithms try to compete with rather low-quality news pieces, often articles produced by news agencies. It seems as if they are doing a good job of that. An increasing number of news outlets is using algorithms for an increasing number of purposes. There are already several companies around the world who provide news algorithm solutions and their solutions will continue to get better. At the same time, producers will find more ways to integrate news algorithms into their daily businesses. It is also just a matter of time until most journalists accept their new “co-workers”. As Wu, Tandoc, & Salmon, 2019, noted, young journalists are already used to working with new technologies like content-generating algorithms.

Secondly, as economic pressure on news companies continues to grow, so will the number of news companies that publish Automated News. News algorithms are seen as a way of relief from political and economic pressure. They help cut down costs and free up journalists from

doing routine work (ibid.). It is therefore safe to assume that news generating algorithms are here to stay.

Finally, as evidenced by the case of the two weather reports, journalists can still distinguish their texts from Automated News. In the case of the weather report, the human journalist simply did not follow the usual way weather reports are written. Participants rated the report as more entertaining, more interesting and better-written. This also caused an effect on the cognitive elaboration. However, human journalists have to be careful when doing that. They are already expected to be more biased and less credible as news algorithms (Haim & Graefe, 2017; Waddell , 2018; Liu & Wei, 2018; Waddell , 2019).

6 Limitations

Obviously, this study had to deal with some limitations. Firstly, the sample of people who participated in the study was not a random sample. Participants were recruited through a snowball system amongst the author's fellow students, through a survey sharing platform also primarily used by students and they were personally recruited on the campus of the University of Vienna. This shows in the sample. The mean age was 24.4, almost 60 percent were students. The gender distribution in the sample (65.5 % females) also does not mirror the distribution among the entire population.

The even larger limitation was the selection of articles. As already mentioned in chapter 5, authors who conduct experiments to compare the different perception between computer-written and human-written articles usually hand-pick their stimulus material. However, as already described, there is a large number of Automated News articles produced every single day. So, hand-picking a low number of them (in most cases two) does not seem like a good idea in hindsight. Future studies are well advised to change that and use a lot more articles as stimulus material. More ideas on that are presented in chapter 7.

It must be mentioned that the experiment was held in English in a German-speaking country. The effects perceived by the participants could be entirely different in a sample where the experiment is held in the native language of most of the participants.

Finally, it is not a good sign, that the results in the two groups, one that rated only one of the two articles on one topic, and the other that rated both articles, differs. It was a good idea to distinguish between the two groups when evaluating the results. Authors of future studies might want to keep in mind, that those two scenarios have an influence on the result and decide for themselves, which experimental design to go with.

7 Conclusion and Outlook

All in all, this study tried to contribute to the better understanding of the different perceptions of computer-written and human-written news. It did so by focusing only on the content without declaring the source of the articles. Interestingly, the general rule, that computer-written news is seen as more credible, whereas human-written news is perceived as more readable, was not entirely confirmed in this setting. This comes as no surprise as it was hypothesized, that the topic would have an influence on the difference in the different perception. Previous findings, that the general rule does not hold true for finance articles, were supported. The more interesting finding is though, that the rule holds true and the effects are large when two rather different articles are used as stimulus material. This difference caused large differences in terms of credibility and readability perceived and in the cognitive elaboration of the participants.

Future work should keep in mind that one single article to represent all computer-written or human-written articles is certainly not enough as one individual article can change the result dramatically. So far, authors hand-picked their stimulus material and presented a very low number of different articles to a comparatively huge sample of persons. It might be an interesting idea to turn that around and select a high number of articles that is more representative of the entirety of Automated News and human-written articles and present them to a limited number of participants. As the content seems to have an influence, it might also be a good idea to analyze the content of Automated News, something that has not been done to the author's knowledge.

We are just at the beginning of the rise of Automated News. It is safe to assume that the number of news articles generated by algorithms will increase dramatically. Automated News will continue to offer a high number of different possibilities for research. And they should be

researched as they are changing journalism, one of the most important institutions in democracies.

Disclosure

The author did not report any conflicts of interest.

Funding

The author received no financial support for the research, authorship, and/or publication of this paper. This paper was written as the author's master's thesis mentored by Univ.-Prof. Dr. Homero Gil de Zúñiga, PhD at the University of Vienna.

References

Clerwall, C. (2014). Enter the Robot Journalist. *Journalism Practice, 8*(5), S. 519-531. doi:10.1080/17512786.2014.883116

Coddington, M. (2015). Clarifying Journalism's Quantitative Turn: A Typology for Evaluating Data Journalism, Computational Journalism, and Computer-Assisted Reporting. *Digital Journalism, 3*(3), S. 331-348. doi:10.1080/21670811.2014.976400

Diakopoulos, N., & Koliska, M. (2017). Algorithmic Transparency in the News Media. *Digital Journalism, 5*(7), S. 809-828. doi:10.1080/21670811.2016.1208053

Dörr, K. (2016). Mapping the field of Algorithmic Journalism. *Digital Journalism, 4*(6), S. 700-722. doi:10.1080/21670811.2015.1096748

Dörr, K., & Hollnuchner, K. (2017). Ethical Challenges of Algorithmic Journalism. *Digital Journalism, 5*(4), S. 404-419. doi:10.1080/21670811.2016.1167612

Eveland Jr., W. (2001). The cognitive mediation model of learning from the news: Evidence from nonelection, off-year election, and Presidential election contexts. *Communication Research,, 28*(5), S. 571-601. doi:10.1177/00936500102800500

Gil de Zúñiga, H. (2017). Attributes of Interpersonal Political Discussion as Antecedents of Cognitive Elaboration. *Revista Española de Investigaciones Sociológicas, 157*, S. 65-84. doi:http://dx.doi.org/10.5477/cis/reis.157.65

Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. (2016). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism*, S. 1-16. doi:10.1177/1464884916641269

Gynnild, A. (2014). The Robot Eye Witness. *Digital Journalism, 2*(3), S. 334-343. doi:10.1080/21670811.2014.883184.

Haim, M., & Graefe, A. (2017). Automated News. *Digital Journalism, 5*(8), S. 1044-1059. doi:10.1080/21670811.2017.1345643

Hoy, M. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly, 37*(1), S. 81-88. doi:10.1080/02763869.2018.1404391

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*, S. 291-298. doi:https://doi.org/10.1016/j.chb.2017.02.022.

Khurana, D., Koli, A., Khatter, K., & Singh, S. (n.d.). Natural Language Processing: State of the Art, Current Trends and Challenges. Retrieved 07 28, 2019, from https://arxiv.org/ftp/arxiv/papers/1708/1708.05148.pdf

Kim, D., & Kim, S. (2017a). Newspaper companies' determinants in adopting robot journalism. *Technological Forecasting and Social Change, 117*, S. 184-195. doi:https://doi.org/10.1016/j.techfore.2016.12.002

Kim, D., & Kim, S. (2017b). Newspaper journalists' attitudes towards robot journalism. *Telematics and Informatics, 35*, S. 340-357. doi:https://doi.org/10.1016/j.tele.2017.12.009

Lee, N., Kim, K., & Yoon, T. (kein Datum). Implementation of robot journalism by programming custombot using tokenization and custom tagging. *2017 19th International Conference on Advanced Communication Technology (ICACT)*, (S. 566-570).

Lewis, S., Sanders, A., & Carmody, C. (2019). Libel by Algorithm? Automated Journalism and the Threat of Legal Liability. *Journalism & Mass Communication Quarterly, 96*(1), S. 60-81. doi:10.1177/1077699018755983

Linden, C.-G. (2017a). Algorithms for Journalism: The future of news work. *The Journal of Media Innovations, 4*(1), S. 60-76. doi:http://dx.doi.org/10.5617/jmi.v4i1.2420

Linden, C.-G. (2017b). Decades of Automation in the Newsroom. *Digital Journalism, 5*(2), S. 123-140. doi:10.1080/21670811.2016.1160791

Liu, B., & Wei, L. (2018). Machine Authorship In Situ. *Digital Journalism*, S. 1-23. doi:10.1080/21670811.2018.1510740

Lokot, T., & Diakopoulos, N. (2016). News Bots. *Digital Journalism, 4*(6), S. 682-699. doi:10.1080/21670811.2015.1081822

Montal, T., & Reich, Z. (2017). I, Robot. You, Journalist. Who is the Author? *Digital Journalism, 5*(7), S. 829-849. doi:10.1080/21670811.2016.1209083

Reiter, E. (2010). Natural Language Generation. In A. Clark, C. Fox, & S. Lappin, *The Handbook of Computational Linguistics and Natural Language Processing* (S. 574-598). Blackwell Publishing Ltd.

Sundar, S.-S. (1999). Exploring receivers' criteria for perception of print and online news. *Journalism & Mass Communication Quarterly, 76*(2), S. 373-386. doi:10.1177/107769909907600213

Thurman, N., Dörr, K., & Kunert, J. (2017). When Reporters Get Hands-on with Robo-Writing. *Digital Journalism, 5*(10), S. 1240-1259. doi:10.1080/21670811.2017.1289819

Thurman, N., Moeller, J., Helberger, N., & Trilling, D. (2019). My Friends, Editors, Algorithms, and I. *Digital Journalism, 7*(4), S. 447-469. doi:10.1080/21670811.2018.1493936

Van Dalen, A. (2012). The algorithms behind the headlines. *Journalism Practice, 6*(5-6), S. 648-658. doi:10.1080/17512786.2012.667268

Van der Kaa, H., & Krahmer, E. (2014). Journalist versus news consumer: The perceived credibility of machine written news. *Proceedings of the Computation+Journalism.* New York.

Waddell, T. (2018). A Robot Wrote This? *Digital Journalism, 6*(2), S. 236-255. doi:10.1080/21670811.2017.1384319

Waddell, T. (2019). Can an Algorithm Reduce the Perceived Bias of News? Testing the Effect of Machine Attribution on News Readers' Evaluations of Bias, Anthropomorphism, and Credibility. *Journalism & Mass Communication Quarterly, 96*(1), S. 82-100. doi:10.1177/1077699018815891

Wölker, A., & Powell, T. (2018). Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism*, S. 1-18. doi:10.1177/1464884918757072

Wu, S., Tandoc, E. C., & Salmon, C. T. (2019). A Field Analysis of Journalism in the Automation Age: Understanding Journalistic Transformations and Struggles Through Structure and Agency. *Digital Journalism, 7*(4), pp. 428-446. doi:10.1080/21670811.2019.1620112

Zhang, Y., Fang, Y., Wei, K.-K., & He, W. (2013). Cognitive elaboration during wiki use in project teams: An empirical study. *Decision Support Systems, 55*, S. 792-801. doi:10.1016/j.dss.2013.03.004

Zheng, Y., Zhong, B., & Yang, F. (2018). When algorithms meet journalism: The user perception to automated news in a cross-cultural context. *Computers in Human Behavior, 86*, S. 266-275. doi:https://doi.org/10.1016/j.chb.2018.04.046

List of Tables

Appendix A – Articles used as stimulus material

Computer-written Sports article, generated by Hero Sports using data from Sportradar. Published by the Washington Post on May, 16th, 2019. Retrieved from: https://www.washingtonpost.com/sports/nhl/sharks-take-2-1-lead-into-game-4-against-the-blues/2019/05/16/d024a590-780d-11e9-a7bf-c8a43b84ee31_story.html?noredirect=on&utm_term=.e607050b28ef (2019-05-17)

Sharks take 2-1 lead into game 4 against the Blues

San Jose Sharks (46-27-9, second in the Pacific Division during the regular season) vs. St. Louis Blues (45-28-9, third in the Central Division during the regular season)

St. Louis; Friday, 8 p.m. EDT

WESTERN CONFERENCE FINALS: San Jose leads series 2-1

BOTTOM LINE: The San Jose Sharks visit the St. Louis Blues in the Western Conference finals with a 2-1 lead in the series. The teams meet Friday for the seventh time this season. The Sharks won the last matchup 5-4 in overtime. Joe Thornton scored a team-high two goals for the Sharks in the victory.

The Blues have gone 24-15-2 in home games. St. Louis has scored 50 power-play goals, converting on 21.1 percent of chances.

The Sharks are 21-16-4 in road games. San Jose ranks third in the league recording 9.7 points per game, averaging 3.5 goals and 6.2 assists.

TOP PERFORMERS: Ryan O'Reilly leads the Blues with a plus-22 in 82 games played this season. Jaden Schwartz has five goals and three assists over the last 10 games for St. Louis.

Brenden Dillon leads the Sharks with a plus-19 in 82 games played this season. Logan Couture has eight goals and four assists over the last 10 games for San Jose.

DURING THE PLAYOFFS: Sharks: Averaging 3.3 goals, 5.5 assists, 4.3 penalties and 10.9 penalty minutes while giving up 3.2 goals per game with a .897 save percentage.

Blues: Averaging 2.8 goals, 5.0 assists, 2.9 penalties and 6.5 penalty minutes while giving up 2.9 goals per game with a .900 save percentage.

Blues Injuries: Carl Gunnarsson: day to day (lower body), Sammy Blais: out (undisclosed).

Sharks Injuries: Radim Simek: out (lower body).

Human-written sports article, written by Ryan Bristlon of wsn.com. Published on May, 16th, 2019. Retrieved from https://www.wsn.com/nhl/st-louis-blues-vs-san-jose-sharks-game-4-2019 (2019-05-17)

Preview of San Jose Sharks vs St. Louis Blues Game 4

San Jose Sharks (Away)

Head Coach: Peter DeBoer {All-Time 400-313-109 SJ, NJD, FLA}

Forward Timo Meier will be the subject of heavy boos at the Enterprise Center in St. Louis Friday night as it was his hand pass that started the play that ended the game. Beyond that, the Blues should be upset with the loss simply because the Sharks got outplayed in game three.

In game three, the San Jose Sharks were only 47.7 percent on the faceoffs. They also got extremely outhit by the Blues (43-21). They also failed to capitalize on their lone power play opportunity and allowed a power play goal on St. Louis' only attempt. The bright points, however, was the continued jelling of the offense. Joe Thornton had his first multi-goal playoff game and Logan Couture and Erik Karlsson continued to shine. Goaltender Martin Jones also had a strong performance in game three.

Jones made 28 saves and looks to have finally settled down in the crease for this series. Throughout his 17 games played, Jones has a goals against average of 2.89 and a save percentage of .904. He is 10-6 and has one assist.

St. Louis Blues (Home)

Head Coach: Craig Berube {All-Time 113-77-34 STL, PHI}

The Blues will have to put the past behind them and not play with too much emotion if they hope to contain the San Jose offense. Like in game two, a huge bounce-back performance is expected from the gritty St. Louis team. Scoring their first power play goal since their series against Dallas should help the team's confidence heading into game four.

The team's physical yet disciplined play was also a great sign. As mentioned earlier, the Blues managed to outhit the Sharks 43-21 and only took one penalty all game. They'll need to keep that type of composure moving forward.

Goaltender Jordan Binnington will once again start for the St. Louis Blues at home. Binnington allowed five goals on 32 shots in game two. His playoff totals include a 2.67 goals against average and a save percentage of .904 - just like his opposition. Binnington has played more than any other goalie in the postseason with 988 minutes. He also has one assist and also four penalty minutes.

Series recap (San Jose leads 1-0) Game one: 6-3 San Jose

Playoff Performers

Logan Couture scored his 14th of the postseason, tying captain Joe Pavelski's total for 2016 - tying them for most goals in a single Sharks playoff run. Couture has six assists to add to that,

giving him 20 points during the postseason - the highest among all players in the Stanley Cup playoffs. He is plus-8 with five power play points, one short handed point, and one game-winner. He’s taken 59 shots on goal.

Jaden Schwartz finds himself leading the St. Louis Blues and their offense, sitting at eighth in the playoff scoring race. Schwartz has netted nine goals and added four assists. He is a plus-4 with one power play point and two game-winning goals.He also boasts a 62.5 percent success rate in the faceoff circle.

Questions to answer

Can the St. Louis Blues keep the emotions in check after their heartbreaking game three loss or will they get caught making mistakes?

Will the officials play a larger-than-normal factor in this game knowing they will be scrutinized from bell to bell by fans, players, and coaches alike? Will the right calls be made?

Can St. Louis keep the power play momentum going after David Perron broke an 0-for-18 stretch?

Computer-written finance story, generated by Automatic Insights, published by the Associated press using data from Zacks Investment Research. Retrieved from https://www.citynews1130.com/2019/04/23/ebay-1q-earnings-snapshot/ (2019-05-16)

EBay: 1Q Earnings Snapshot

SAN JOSE, Calif. (AP) _ EBay Inc. (EBAY) on Tuesday reported first-quarter earnings of $518 million.

On a per-share basis, the San Jose, California-based company said it had profit of 57 cents. Earnings, adjusted for one-time gains and costs, were 67 cents per share.

The results exceeded Wall Street expectations. The average estimate of 13 analysts surveyed by Zacks Investment Research was for earnings of 63 cents per share.

The e-commerce company posted revenue of $2.64 billion in the period, also surpassing Street forecasts. Ten analysts surveyed by Zacks expected $2.58 billion.

For the current quarter ending in July, eBay expects its per-share earnings to range from 61 cents to 63 cents.

The company said it expects revenue in the range of $2.64 billion to $2.69 billion for the fiscal second quarter. Analysts surveyed by Zacks had expected revenue of $2.66 billion.

EBay expects full-year earnings in the range of $2.64 to $2.70 per share, with revenue ranging from $10.83 billion to $10.93 billion.

EBay shares have increased 30 per cent since the beginning of the year, while the Standard & Poor’s 500 index has increased 17 per cent. In the final minutes of trading on Tuesday, shares hit $36.62, a drop of 12 per cent in the last 12 months.

Human written finance article, written by Claudia Assis of marketwatch.com. Retrieved from https://www.marketwatch.com/story/ebay-stock-rallies-more-than-5-on-quarterly-earnings-beat-increased-guidance-2019-04-23 (2019-05-16)

EBay stock rallies more than 5% on quarterly earnings beat, increased guidance

Shares of EBay Inc. rose more than 5% in the extended session Tuesday after the company reported a first-quarter adjusted profit above Wall Street expectations and raised earnings and sales guidance for the year. EBay said it earned $521 million, or 57 cents a share, in the quarter, compared with $407 million, or 40 cents a share, in the year-ago quarter. Adjusted for one-time items, the company earned $608 million, or 67 cents a share, compared with $548 million, or 53 cents a share, a year ago. Revenue rose 2% to $2.64 billion, compared with $2.58 billion a year ago, the company said.

Analysts polled by FactSet had expected EBay to report adjusted earnings of 63 cents a share on sales of $2.6 billion. EBay saw "healthy buyer growth and disciplined cost control," Chief Executive Devin Wenig said in a statement. The company gained "increased confidence in the year" and adjusted its guidance, Wenig said. EBay expects net revenue between $10.83 billion and $10.93 billion, which would represent growth between 2% and 3%, and non-GAAP earnings between $2.64 a share and $2.70 a share. Shares of EBay ended the regular trading day up 1.4%.

Computer-written news article, provided to the author by e-mail by Retresco, a German company specializing in algorithmic news solutions.

Weather Forecast for the UK

Today

There will be a mix of sun and clouds. A gentle breeze will occur through the day. Daytime temperatures will be 18 to 21°C (64 to 70°F).

This Evening and Tonight

It is going to be dry and clear with patches of cloud in the evening and into the night. There will be squally winds. Temperatures will be 9 to 12°C (48 to 54°F). The night is going to be wet in some areas.

Tomorrow

There are going to be showers or longer spells of rain in parts of the region in the morning and into the afternoon. The day is expected to be mainly cloudy with occasional bright spells. A moderate southwesterly breeze can be expected. Daytime temperatures will be at 19 to 22°C (66 to 72°F).

Outlook for the Next Three Days

It will be raining during the day on Tuesday in places, except for a short period of time in the afternoon. Similar weather can be expected on Wednesday. Thursday should be partially cloudy with sunny spells. Throughout the afternoon, it will be showery.

Transport Forecast

There is a risk of slippery roads in places due to extremely negative RSTs with patches of sleet between 13:00 and 18:00. Sleet and ice but with positive RSTs between 19:00 and 17:00. No risk of slipperiness. Further information can be seen in the table.

Human-written weather report, written by Jenny Awford. Published by The Sun on May, 16th, 2019. Retrieved from https://www.thesun.co.uk/news/9086146/uk-weather-forecast-sun-today-thunderstorms-met-office/ (2019-05-16)

UK weather forecast: Highs of 20C today before thunderstorms hit weekend

BRITS are set to enjoy sunshine and highs of 20C today – but the glorious weather won't last long as thunderstorms could hit this weekend, according to the Met Office.

Weather forecasters warned there could be some showers in the far north and west of both Scotland and Northern Ireland.
But the majority of the country will be dry with sunny skies.

Stornoway in western Scotland could see highs of 20C and Glasgow will enjoy a balmy 19C, whereas London will see temperatures of around 16C.

The glorious blue skies and warm weather won't stay for long, as thunderstorms are expected to roll in over the weekend.

Weather forecasters predict high temperatures will continue until Friday – when thermometers will take a dip once again.

There will be a mix of sunshine and showers throughout the weekend, with some turning thundery and heavy at times.

From Monday forecasters say there is a risk of longer spells of heavy rain and thunderstorms across the south, as the "changeable" weather lasts all of next week.

Met Office weather maps for the week ahead show temperatures reaching the low- to mid-20s – with the area between Glasgow and Edinburgh seeing highs of 23C.

The east of England will be a little cooler with highs of only 14C-16C – but skies should be clear and bright.

The Met Office has also said UV levels are high across the UK this week.

They recommend sunseekers stay in the shade during midday hours as well as covering up and slapping on sunscreen.

Meteorologist Craig Snell told The Sun Online: "We're going to be expecting very dry weather with plenty of sun within the next two days.

"We’re going to see some rain creep in the the second half of the week, as temperatures will begin to settle down on Friday."

The Met Office’s Sophie Yeomans urged people to “make the most” of the sunshine – as temperatures will dip back down later this week.

She said: “We are expecting another couple of days of sunny weather and light winds, which will push temperatures up through the week."

Appendix B – More detailed tables on the results

Table 7

More detailed comparison of the credibility attributes between participants who rated only one sports article

Attribute	Computer-written			Human-written			Test statistics				
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Mann-Whitney-U	Wilcoxon-W	Z	p	r
Accurate	127	137.07	17408.50	122	112.43	13716.50	6213.50	13716.50	-2.796	0.005	0.18*
Trustworthy	127	131.03	16641.00	123	119.97	14734.00	7108.00	14734.00	-1.267	0.205	-
Fair	126	131.19	16530.50	122	117.59	14345.50	6842.50	14345.50	-1.555	0.120	-
Reliable	124	133.14	16509.00	122	113.70	13872.00	6369.00	13872.00	-2.203	0.028	0.14*

Note: * Small effect (0.1 < r < 0.3). ** Medium effect (0.3 < r < 0.5). *** Large effect (r > 0.5).

Table 8

More detailed comparison of the readability attributes between participants who rated only one sports article

Attribute	Computer-written			Human-written			Test statistics				
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Mann-Whitney-U	Wilcoxon-W	Z	p	r
Entertaining	127	120.41	15292.50	123	130.75	16082.50	7164.50	15292.50	-1.148	0.251	-
Interesting	127	123.03	15624.50	123	128.05	15750.50	7496.50	15624.50	-0.559	0.576	-
Vivid	126	110.15	13989.00	122	138.66	16639.00	5861.00	13989.00	-3.204	0.001	0.20*
Well-written	124	112.94	14344.00	122	138.46	17031.00	6216.00	14344.00	-2.854	0.004	0.18*

Note: * Small effect (0.1 < r < 0.3). ** Medium effect (0.3 < r < 0.5). *** Large effect (r > 0.5).

Table 9

More detailed comparison of the cognitive elaboration attributes between participants who rated only one sports article

Attribute	Computer-written			Human-written			Test statistics				
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Mann-Whitney-U	Wilcoxon-W	Z	p	r
I thought about the content of this article.	127	125.57	15947.00	123	125.42	15428.00	7802.00	15428.00	-0.015	0.988	-
I thought about similar news articles I read in the past.	127	133.32	16931.50	123	117.43	14443.50	6817.50	14443.50	-1.789	0.074	-
I tried to connect the content of this article to my knowledge about the topic.	127	122.09	15505.00	122	128.03	15620.00	7377.00	15505.00	-0.660	0.509	-
The content of this article is interesting to me.	127	124.26	15781.00	123	126.78	15594.00	7653.00	15781.00	-0.285	0.776	-

Note: * Small effect (0.1 < r < 0.3). ** Medium effect (0.3 < r < 0.5). *** Large effect (r > 0.5).

Table 10

More detailed comparison of the credibility attributes between participants who rated only one economics article

Attribute	Computer-written			Human-written			Test statistics				
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Mann-Whitney-U	Wilcoxon-W	Z	p	r
Accurate	122	114.39	13955.00	129	136.98	17671.00	6452.00	13955.00	-2.546	0.011	0.16*
Trustworthy	123	116.27	14301.50	128	135.35	17324.50	6675.50	14601.50	-2.136	0.033	0.13*
Fair	123	112.37	13822.00	128	139.09	17804.00	6196.00	13822.00	-3.058	0.002	0.19*
Reliable	123	120.00	14760.00	130	133.62	17371.00	7134.00	14760.00	-1.526	0.127	-

Note: * Small effect (0.1 < r < 0.3). ** Medium effect (0.3 < r < 0.5). *** Large effect (r > 0.5).

Table 11

More detailed comparison of the readability attributes between participants who rated only one finance article

Attribute	Computer-written			Human-written			Test statistics				
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Mann-Whitney-U	Wilcoxon-W	Z	p	r
Entertaining	123	122.04	15011.00	130	131.69	17120.00	7385.00	15011.00	-1.074	0.283	-
Interesting	123	123.36	15173.50	129	129.49	16704.50	7547.50	15173.50	-0.678	0.498	-
Vivid	121	115.22	13942.00	125	131.51	16439.00	6561.00	13942.00	-1.841	0.066	0.12*
Well-written	123	115.83	14247.00	129	136.67	17631.00	6621.00	14247.00	-2.315	0.021	0.15*

Note: * Small effect (0.1 < r < 0.3). ** Medium effect (0.3 < r < 0.5). *** Large effect (r > 0.5).

Table 12

More detailed comparison of the cognitive elaboration attributes between participants who rated only one finance article

Attribute	Computer-written			Human-written			Test statistics				
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Mann-Whitney-U	Wilcoxon-W	Z	p	r
I thought about the content of this article.	123	122.54	15072.50	128	129.32	16533.50	7446.50	15072.50	-0.750	0.453	-
I thought about similar news articles I read in the past.	123	122.36	15050.50	129	130.45	16827.50	7424.50	15050.50	-0.895	0.371	-
I tried to connect the content of this article to my knowledge about the topic.	123	123.74	15220.00	129	129.13	16658.00	7594.00	15220.00	-0.596	0.551	-
The content of this article is interesting to me.	123	122.83	15107.50	128	129.05	16518.50	7481.50	15107.50	-0.690	0.490	-

Note: * Small effect (0.1 < r < 0.3). ** Medium effect (0.3 < r < 0.5). *** Large effect (r > 0.5).

Table 13

More detailed comparison of the credibility attributes between participants who rated only one weather article

Attribute	Computer-written			Human-written			Test statistics				
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Mann-Whitney-U	Wilcoxon-W	Z	p	r
Accurate	136	109.50	14892.00	114	144.59	16483.00	5576.00	14892.00	-4.622	0.000	0.29*
Trustworthy	136	159.50	21692.00	114	84.94	9683.00	3128.00	9683.00	-9.408	0.000	0.59***
Fair	136	131.50	17884.00	114	118.34	13491.00	6936.00	13491.00	-1.697	0.090	0.11*
Reliable	123	120.00	14760.00	114	78.38	8935.00	2380.00	8935.00	-11.124	0.000	0.70***

Note: * Small effect (0.1 < r < 0.3). ** Medium effect (0.3 < r < 0.5). *** Large effect (r > 0.5).

Table 14

More detailed comparison of the readability attributes between participants who rated only one weather article

Attribute	Computer-written			Human-written			Test statistics				
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Mann-Whitney-U	Wilcoxon-W	Z	p	r
Entertaining	136	75.50	10268.00	114	185.15	21107.00	952.00	10268.00	-13.511	0.000	0.85***
Interesting	136	107.50	14620.00	114	146.97	16755.00	5304.00	14620.00	-4.948	0.000	0.31**
Vivid	136	119.50	16252.00	113	131.62	14873.00	6936.00	16252.00	-1.603	0.109	-
Well-written	136	144.50	19652.00	114	102.83	11723.00	5168.00	11723.00	-5.290	0.000	0.33**

Note: * Small effect (0.1 < r < 0.3). ** Medium effect (0.3 < r < 0.5). *** Large effect (r > 0.5).

Table 15

More detailed comparison of the cognitive elaboration attributes between participants who rated only one weather article

Attribute	Computer-written			Human-written			Test statistics				
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Mann-Whitney-U	Wilcoxon-W	Z	p	r
I thought about the content of this article.	136	94.50	12852.00	113	161.71	18273.00	3536.00	12852.00	-8.265	0.000	0.52***
I thought about similar news articles I read in the past.	136	128.00	17408.00	113	121.39	13717.00	7276.00	13717.00	-0.840	0.401	-
I tried to connect the content of this article to my knowledge about the topic.	136	147.00	19992.00	113	98.52	11133.00	4692.00	11133.00	-6.008	0.000	0.38**
The content of this article is interesting to me.	136	114.00	15504.00	113	138.24	15621.00	6188.00	15504.00	-3.079	0.002	0.20*

Note: * Small effect (0.1 < r < 0.3). ** Medium effect (0.3 < r < 0.5). *** Large effect (r > 0.5).

Table 16

More detailed comparison of the credibility attributes between participants who rated both sports articles

Attribute	Negative Ranks			Positive Ranks					Test statistics		
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Connections	Total	Z	p	r
Accurate	25	20.04	501.00	16	22.50	360.00	38	79	-0.929	0.353	-
Trustworthy	28	21.32	597.00	18	26.89	484.00	32	78	-0.631	0.528	-
Fair	30	22.92	687.50	13	19.98	258.50	35	78	-2.627	0.009	0.30**
Reliable	31	21.97	687.00	14	25.29	354.00	32	77	-1.894	0.058	0.21*

Note: * Small effect (0.1 < r < 0.25). ** Medium effect (0.25 < r < 0.4). *** Large effect (r > 0.4).

Table 17

More detailed comparison of the readability attributes between participants who rated both sports articles

Attribute	Negative Ranks			Positive Ranks				Test statistics		
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Connections	Total	Z	p	r
Entertaining	23	22.96	528.00	28	28.50	798.00	26	77	-1.280	0.200	-
Interesting	21	21.48	451.00	26	26.04	677.00	31	78	-1.215	0.224	-
Vivid	17	20.47	348.00	35	29.43	1030.00	24	76	-3.135	0.002	0.35**
Well-written	19	25.61	486.50	35	28.53	998.50	24	78	-2.265	0.025	0.26**

Note: * Small effect (0.1 < r < 0.25). ** Medium effect (0.25 < r < 0.4). *** Large effect (r > 0.4).

Table 18

More detailed comparison of the cognitive elaboration attributes between participants who rated both sports articles

Attribute	Negative Ranks			Positive Ranks					Test statistics		
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Connections	Total	Z	p	r
I thought about the content of this article.	17	27.26	463.50	32	23.80	761.50	28	77	-1.532	0.126	-
I thought about similar news articles I read in the past.	21	22.69	476.50	19	18.08	343.50	38	78	-0.905	0.365	-
I tried to connect the content of this article to my knowledge about the topic.	22	24.09	530.00	27	25.75	695.00	27	76	-0.843	0.399	-
The content of this article is interesting to me.	15	16.10	241.50	18	17.75	319.50	45	78	-0.714	0.475	-

Note: * Small effect (0.1 < r < 0.25). ** Medium effect (0.25 < r < 0.4). *** Large effect (r > 0.4).

Table 19

More detailed comparison of the credibility attributes between participants who rated both finance articles

Attribute	Negative Ranks			Positive Ranks					Test statistics		
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Connections	Total	Z	p	r
Accurate	15	17.03	255.50	20	18.73	374.50	33	68	-1.027	0.305	-
Trustworthy	17	18.32	311.50	20	19.58	391.50	32	69	-0.625	0.532	-
Fair	18	17.61	317.00	16	17.38	278.00	34	68	-0.353	0.724	-
Reliable	18	17.64	317.50	16	17.34	277.50	35	69	-0.354	0.723	-

Note: * Small effect (0.1 < r < 0.25). ** Medium effect (0.25 < r < 0.4). *** Large effect (r > 0.4).

Table 20

More detailed comparison of the readability attributes between participants who rated both business articles

Attribute	Negative Ranks			Positive Ranks					Test statistics		
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Connections	Total	Z	p	r
Entertaining	23	21.09	485.00	22	25.00	550.00	22	68	-0.374	0.708	-
Interesting	15	18.47	277.00	24	20.96	503.00	29	68	-1.626	0.104	-
Vivid	20	22.65	453.00	23	21.43	493.00	26	69	-0.246	0.806	-
Well-written	24	18.50	444.00	23	29.74	684.00	22	69	-1.284	0.199	-

Note: * Small effect (0.1 < r < 0.25). ** Medium effect (0.25 < r < 0.4). *** Large effect (r > 0.4).

Table 21

More detailed comparison of the cognitive elaboration attributes between participants who rated both business articles

Attribute	Negative Ranks			Positive Ranks					Test statistics		
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Connections	Total	Z	p	r
I thought about the content of this article.	16	15.53	248.50	16	17.47	279.50	37	69	-0.297	0.766	-
I thought about similar news articles I read in the past.	20	18.63	372.50	17	19.44	330.50	32	69	-0.322	0.747	-
I tried to connect the content of this article to my knowledge about the topic.	20	19.02	380.50	18	20.03	360.50	31	69	-0.148	0.882	-
The content of this article is interesting to me.	15	16.40	246.00	21	20.00	420.00	33	69	-1.426	0.154	-

Note: * Small effect (0.1 < r < 0.25). ** Medium effect (0.25 < r < 0.4). *** Large effect (r > 0.4).

Table 22

More detailed comparison of the credibility attributes between participants who rated both weather articles

Attribute	Negative Ranks			Positive Ranks					Test statistics	
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Connections	Total	Z	p	r
Accurate	37	32.30	1195.00	23	27.61	635.00	32	92	-2.112	0.035	0.22*
Trustworthy	41	29.06	1191.50	17	30.56	519.50	35	92	-2.665	0.008	0.28**
Fair	34	27.46	933.50	25	33.46	836.50	34	92	-0.376	0.707	-
Reliable	43	30.98	1332.00	19	32.68	621.00	30	92	-2.580	0.010	0.27**

Note: * Small effect (0.1 < r < 0.25). ** Medium effect (0.25 < r < 0.4). *** Large effect (r > 0.4).

Table 23

More detailed comparison of the readability attributes between participants who rated both weather articles

Attribute	Negative Ranks			Positive Ranks					Test statistics		
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Connections	Total	Z	p	r
Entertaining	29	37.36	1083.50	44	36.76	1617.50	19	92	-1.484	0.138	-
Interesting	41	33.93	1391.00	26	34.12	887.00	25	92	-1.601	0.109	-
Vivid	26	28.31	736.00	42	38.33	1610.00	22	90	-2.701	0.007	0.28**
Well-written	41	37.95	1556.00	31	34.58	1072.00	20	92	-1.387	0.166	-

Note: * Small effect (0.1 < r < 0.25). ** Medium effect (0.25 < r < 0.4). *** Large effect (r > 0.4).

Table 24

More detailed comparison of the cognitive elaboration attributes between participants who rated both weather articles

Attribute	Negative Ranks			Positive Ranks					Test statistics		
	N	Mean Rank	Rank sum	N	Mean Rank	Rank sum	Connections	Total	Z	p	r
I thought about the content of this article.	32	28.14	900.50	29	34.16	990.50	31	92	-0.337	0.736	-
I thought about similar news articles I read in the past.	33	29.32	967.50	21	24.64	517.50	38	92	-1.971	0.049	0.21*
I tried to connect the content of this article to my knowledge about the topic.	25	24.02	600.50	27	28.80	777.50	40	92	-0.831	0.406	-
The content of this article is interesting to me.	34	30.49	1036.50	24	28.10	647.50	34	92	-1.455	0.146	-

Note: * Small effect (0.1 < r < 0.25). ** Medium effect (0.25 < r < 0.4). *** Large effect (r > 0.4).

Abstract

Automated News, news that are produced automatically by algorithms that analyze data, are increasingly popular among news companies. Those algorithms analyze data and produce journalistic articles much faster and cheaper than human journalists. Obviously, the long-term success of this new form of journalism depends on the perception of the audience, which is the main topic of this study. Previous studies found out that in general computer-written news tend to be considered more credible, whereas human-written news is considered more readable. This general rule does not always apply and depends on some factors, as past studies have shown. This paper is dedicated to finding out, if the topic is one of those factors, as previous research indicated. The paper discusses, how the rule does not seem to apply for finance articles. It also shows that the measured differences are also heavily influenced by the individual articles that are chosen as stimulus material.

Algorithmische Nachrichten, Nachrichten, die von Algorithmen, die Daten analysieren, automatisch produziert werden, werden immer beliebter in Nachrichtenhäusern. Die Algorithmen analysieren Daten und produzieren Artikel schneller und billiger als Menschen es können. Klarerweise hängt der langfristige Erfolg dieser neuen Form des Journalismus von der Wahrnehmung des Publikums ab. Sie ist das zentrale Thema dieser Arbeit. In vergangenen Arbeiten wurde die generelle Regel aufgestellt, dass von Computern erzeugte Nachrichten glaubhafter wirken, während von Menschen geschriebene Nachrichten lesenswerter sind. Diese allgemeine trifft allerdings nicht immer zu und hängt von verschiedenen Faktoren ab, wie in vergangenen Studien gezeigt wurde. Diese Arbeit widmet sich der Frage, ob das Thema des Nachrichtenartikels einer dieser Faktoren ist, wie es in vergangenen Studien vermutet wurde. Es wird festgestellt, dass die Regel bei Artikeln über Finanzen nicht zutreffen dürfte. Außerdem wird gezeigt, dass die gemessenen Unterschiede stark von der Auswahl der individuellen Artikel, die in den verschiedenen Studien als Stimulus Material verwenden werden, abhängen.