PhDr. Václav Moravec, Ph.D., Ph.D.
Centre of Artificial Intelligence Journalism
Institute of Communication Studies and Journalism
Faculty of Social Sciences
Charles University
Smetanovo nábřeží 6
110 01 Prague 1
Czech Republic
vaclav.moravec@fsv.cuni.cz

Václav Moravec works at the Institute of Communication Studies and Journalism at the Faculty of Social Sciences, Charles University in Prague and at the Department of Production at FAMU. He specialises in transformation of audio-visual media, journalistic ethics, automated journalism and artificial intelligence journalism. He has been working for *Czech Television* in the TV discussion programmes *Otázky Václava Moravce* (since 2004) and *Fokus Václava Moravce* (since 2015). He is also a member of the Executive Committee of the Artificial Intelligence Initiative PRG.AI.

Mgr. Veronika Macková
Centre of Artificial Intelligence Journalism
Institute of Communication Studies and Journalism
Faculty of Social Sciences
Charles University
Smetanovo nábřeží 6
110 01 Prague 1
Czech Republic
veronika.mackova@fsv.cuni.cz

Veronika Macková is a Ph.D. candidate, researcher and assistant lecturer at the Institute of Communication Studies and Journalism at the Faculty of Social Sciences, Charles University in Prague. In her academic research, she focuses on issues associated with sport and disabilities, artificial intelligence and journalism, solutions journalism and cross-platform media. She gives lectures related to practical journalism and the history of sports journalism at the Faculty of Social Sciences of Charles University. She also works for Czech public television as a sports reporter.
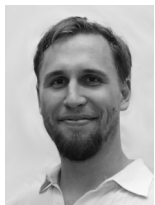
Ing. Jakub Sido
Department of Computer Science and Engineering
Faculty of Applied Sciences
University of West Bohemia
Univerzitní 8
301 00 Pilsen
Czech Republic
sidoj@kiv.zcu.cz

Jakub Sido is a Ph.D. student and researcher at the Department of Computer Science and Engineering at the Faculty of Applied Sciences, University of West Bohemia, Czech Republic. In his research, he focuses on artificial intelligence and natural language processing and generation.

Ing. Kamil Ekštein, Ph.D.
Department of Computer Science and Engineering
Faculty of Applied Sciences
University of West Bohemia
Univerzitní 8
301 00 Pilsen
Czech Republic
kekstein@kiv.zcu.cz

Kamil Ekštein works as an Assistant Professor and researcher at the Department of Computer Science and Engineering of the University of West Bohemia. He obtained his Ph.D. with chancellor's appreciation for outstanding research in 2004 from the UWB, specialising in hybrid methods of spontaneous speech analysis. His professional interests are focused on general artificial intelligence and machine learning, digital signal processing, and speech and image recognition. He teaches a graduate course on machine learning. He also collaborates with many academic and commercial research facilities as a lecturer and consultant, both in the Czech Republic and abroad.

THE ROBOTIC REPORTER IN THE CZECH NEWS AGENCY: AUTOMATED JOURNALISM AND AUGMENTATION IN THE NEWSROOM

Václav MORAVEC – Veronika MACKOVÁ – Jakub SIDO – Kamil EKŠTEIN

ABSTRACT:
The use of automated journalism, also known as robotic journalism or artificial intelligence journalism, became an established practice in English-speaking countries less than ten years ago. Narrative Science and Automated Insights developed creative software that automatically generates reports. Several media outlets, including *The Associated Press* (AP), have started to publish their reports. Media landscape barriers based on Slavic languages, such as Czech, have caused some delays in the introduction of automated journalism, or artificial intelligence journalism, in Central and Eastern Europe. This article is a case study of the application of algorithms that transform large data files into news texts in *The Czech News Agency* (ČTK). A research team led by Charles University provided algorithms generating reports on trading results on the Prague Stock Exchange without human intervention to *The Czech News Agency* in 2019. The study deals with the production of algorithms and compares the rate of generation of messages generated by humans against algorithms and examines their quality. In our research we also used observations and questionnaire surveys of selected journalists and editors who work with reports from the Prague Stock Exchange. The article also provides the opinions of journalists of *The Czech News Agency* on the application of automated journalism and artificial intelligence journalism in their newsrooms.

KEY WORDS:
algorithmic journalism, automated journalism, big data, Czech News Agency, journalistic ethics, natural language generation, natural language processing, news forms, Prague Stock Exchange, robotic journalism

Introduction

In the afternoon of 1st November 2019, when trading on the Prague Stock Exchange ended, the algorithms generated a report in Czech about the results of the trading day in one-tenth of a second. The news text of two hundred words was similar to those published every weekday by *The Czech News Agency* on its service within

an hour of the closing of the stock exchange. The software marked a significant advance in the application of automated, or robotic journalism, in the Czech Republic. Scientists have given the software the nickname *R.U.R.* in reference to a play by the Czech journalist and writer Karel Čapek, who originated the word "robot" which then spread to all languages of the world.[1] The adoption of this type of journalism, which reflects the latest trends, has been slower in the Czech media landscape – not only because of the linguistically limited market of the 10 million person nation (including the limited number of news texts of individual thematic domains to which machine learning elements could be applied) but also because of the relatively high costs of developing similar creative software, for which there is no money left in their budgets due to the underfunding of the Czech media.

The software generates reports on trading results based on data from the Prague Stock Exchange every weekday. These reports are forwarded to the Digital Media Systems publishing programme (services enabled to create, manage and distribute content) of *The Czech News Agency* and *Czech Television*. Czech public service media (*The Czech News Agency* and *Czech Television*) try to rank as leading news organisations in English-speaking countries such as *Associated Press*, *Reuters* or *The New York Times*. They have elements of artificial intelligence as part of their daily editorial practice.[2]

This article is a case study of the first experiences with automated journalism, which we understand as sophisticated algorithmic processes transforming large data into news stories without human intervention. It builds on the comparison of the reports on trading results on the Prague Stock Exchange which were written by journalists and, on the other side, by a robotic 'reporter'. It brings a reflection on the first experience of the management of *The Czech News Agency* and the news agency editors and journalists with the software *R.U.R.* The software was invented through the cooperative effort of three academic institutions – the Faculty of Social Sciences, Charles University in Prague; the Faculty of Electrical Engineering, Czech Technical University in Prague; and the Faculty of Applied Sciences in Pilsen. This reflection is based on questionnaire surveys and structured interviews with the three senior editors of *The Czech News Agency* and seven reporters (employees) of its economic editorial office, who are involved in the creation of stock market reporting, and participating observation for three days in *The Czech News Agency*. At the beginning of the development of automated journalism in the Czech Republic, we have the following research questions: What is the change in the production and presentation of reports? How will elements of automation and artificial intelligence change editorial practices regarding public service media?

Artificial Intelligence in Journalism – The Problem of a Definitional Framework

The concept of "artificial intelligence" (AI) has become fashionable in many social sciences – including journalism and media studies, which distracts attention from the true nature of its content. In the phrase "artificial intelligence", the public imagines machines or beings that compete with human abilities to learn, understand in context and be creative. In these cases, the appropriate term to use is "artificial general intelligence" (AGI).[3] We know them not only from Čapek's drama, *R.U.R.*, but also from the work of the British writer Arthur C. Clarke, *2001: A Space Odyssey*, where the onboard supercomputer, HAL 9000, communicates and collaborates with the crew of the Discovery One.[4] The perception of the terms AI and AGI as synonyms contributes to the problematic conceptual framework around artificial intelligence because AGI is still just another hypothesis. Artificial intelligence is still many years from being anything other than

theoretical.[5] Computers are faster than humans, but in reality, this is "machine learning" (ML), "natural language processing" (NLP)[6] or "natural language generation" (NLG)[7] but they were all created by humans.

Artificial intelligence or automated journalism are currently used in several areas in journalism – for collecting information (structured data, processing unstructured information), verifying information, creating texts, visualisations and graphics, and for targeted distribution of information. Many media organisations have used machine learning techniques to solve various tasks.[8] It was the automatic generation of news texts that led to the notion of artificial intelligence in journalism a decade ago. From a theoretical background, researchers need to focus on areas such as text linguistics, computational narrative and knowledge engineering.[9] Claims are increasing for journalists, who can now use data analysis or programming skills.[10] It will therefore be *"a combination of algorithms, data, and knowledge from the social sciences to supplement the accountability function of journalism".*[11]

The software company Narrative Science and its creative system The Quill started to generate brief news summaries for baseball and softball matches for the *American Big Ten Network* (BTN) in spring 2010 without human intervention.[12] *BTN.com* published them one or two minutes after the end of each match, leading to a significant increase in traffic to the website.[13] According to Caswell and Dörr, platforms such as Narrative Science's Quill can create templates: *"Each template is designed to operate with a particular data model, and can then be used by the platform to assemble blocks of written text from its text fragments based on any dataset with fields formatted in accordance with that data model."*[14] Another example is the prominent software company, Automated Insights, which is referred to as the second major player in this field. Its software, Wordsmith, has written reports for the *Associated Press* (AP) on quarterly financial results of US listed companies since 2014.[15] Elements of artificial intelligence increased the number of published articles in AP's service twelve times (to almost four thousand), and much more attention began to be paid to smaller companies, whose AP's quarterly results were overlooked before the advent of robotic journalism due to capacity reasons.[16] In Figure 1 is listed the percentage development of AP reports on US business results. Journalists reported only about 30% of business results in 2012. This fact was changed due to the advancement of artificial intelligence, which in 2014 managed to process 70% of business results. The sample included 4,292 company and 57,467 economic results reports.[17] Blankespoor, DeHaan and Zhu found that in this case *"automated articles increase firms' trading volume and liquidity. The effects are most likely driven by retail traders".*[18]

1   ČAPEK, K.: *R.U.R.* Prague : Artur, 2004, p. 102.
2   MORAVEC, V.: *Proměny novinářské etiky*. Prague : Academia, 2020, p. 83-85.
3   FRANKLIN, S.: A Foundational Architecture for Artificial General Intelligence. In GOERTZEL, B., WANG, P. (eds.): *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms*. Amsterdam : IOS Press, 2007, p. 36-54.
4   CLARKE, A. C.: *2001: Vesmírná odysea*. Prague : Laser, 2008.

5   BODEN, M.: Artificial Intelligence. In AL-KHALILI, J. (ed.): *What's Next?* London : Profile Books, 2017, p. 118-128.
6   MCDONALD, D. D.: Natural Language Generation. In INDURKHYA, N., DAMERAU, F. J.: *Handbook of Natural Language Processing*. Boca Raton : Chapman and Hall, 2010, p. 121-141.
7   DEEMTER, K., KRAHMER, E., THEUNE, M.: Real versus Template-Based Natural Language Generation: A False Opposition? In *Computational Linguistics*, 2005, Vol. 31, No. 1, p. 15-24.
8   STRAY, J.: Making Artificial Intelligence Work for Investigative Journalism. In *Digital Journalism*, 2019, Vol. 7, No. 8, p. 1076-1097.
9   CASWELL, D.: Structured Journalism and the Semantic Units of News. In *Digital Journalism*, 2019, Vol. 7, No. 8, p. 1134-1156.
10   See: MAYER-SCHÖNBERGER, V., CUKIER, K.: *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London : John Murray Publishers, 2013.
11   HAMILTON, J. T., TURNER, F.: *Accountability Through Algorithm: Developing the Field of Computational Journalism*. [online]. [2020-03-08] Available at: <https://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>.
12   MARR, B.: *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*. London : Wiley, 2016.
13   MORAVEC, V.: *Proměny novinářské etiky*. Prague : Academia, 2020, p. 82.
14   CASWELL, D., DÖRR, K.: Automated Journalism 2.0: Event-Driven Narratives. In *Journalism Practice*, 2018, Vol. 12, No. 4, p. 477-496.
15   GRAEFE, A.: *Guide to Automated Journalism*. New York : Tow Center for Digital Journalism, Columbia Journalism School, 2016. [online]. [2019-12-02]. Available at: <https://academiccommons.columbia.edu/doi/10.7916/D80G3XDJ>.
16   MARCONI, F., SIEGMAN, A.: *The Future of Augmented Journalism: A Guide for Newsrooms in the Age of Smart Machines*. New York : Associated Press, 2017, p. 4. [online]. [2019-12-02]. Available at: <https://insights.ap.org/uploads/images/the-future-of-augmented-journalism_ap-report.pdf>.
17   *Annual Report 2016*. [online]. [2020-03-07]. Available at: <https://automatedinsights.com>.
18   BLANKESPOOR, E., DEHAAN, E., ZHU, C.: Capital Market Effects of Media Synthesis and Dissemination: Evidence from Robo-journalism. In *Review of Accounting Studies*, 2018, Vol. 23, No. 1, p. 1.
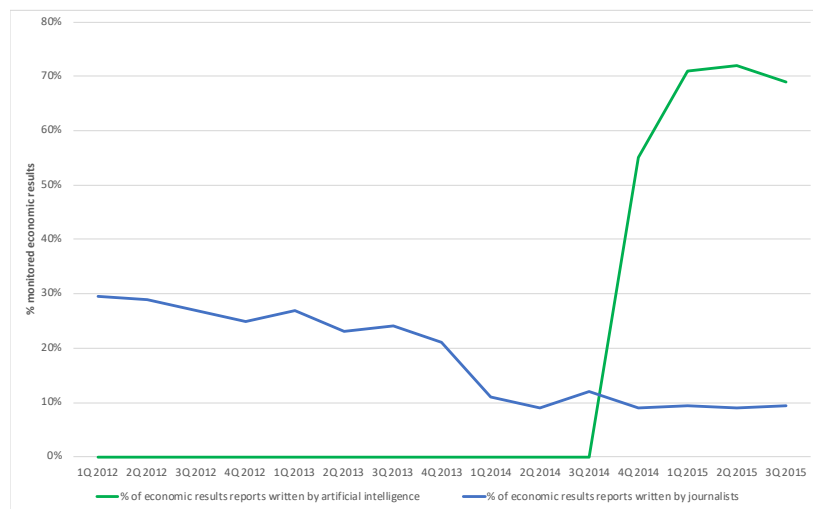
Figure 1: Percentage development of Associated Press (AP) reports on quarterly US financial results written by journalists and artificial intelligence

Source: Annual Report 2016. [online]. [2020-03-07]. Available at: <https://automatedinsights.com>.

Dozens of creative software tools that work on machine learning, natural language processing or natural language generation techniques and use structured data sources, produce thousands of stories each month for major news organisations in many thematic areas – including sports, economics, politics and weather (such as *BBC, Reuters, The New York Times*, etc.).[19] *The New York Times* publishes an application for news aggregation and machine learning which is also used by *Google News*. Furthermore, automated journalism is now being used as well by *The Czech News Agency* and by *Czech Television*.

Terms such as "robotic journalism" or "automatic journalism" appear more and more frequently.[20] Reflection of the first successes of automatically generated and published reports led to the defining of a new section of journalism as "robotic"[21] or "automated".[22] Automated journalism or robotic journalism as synonymous terms include the innovative use of algorithms, artificial intelligence software platforms, and natural language generation techniques. Artificial intelligence journalism has subsequently become an umbrella term.[23] Automated journalism includes algorithmic journalism, which Dörr defines as *"the (semi) automated process of natural language generation by the selection of electronic data from private or public databases (input), the assignment of relevance to pre-selected or non-selected data characteristics, processing and structuring of the relevant data sets to a semantic structure (throughput), and publishing of the final text on an online or offline platform with a certain reach (output)".*[24]

AI technologies' range develops from everyday functions and evolves rapidly. Using artificial intelligence is not as dramatic as starting to use the Internet. What the final impact of artificial intelligence journalism will be is still unknown, but there is great potential for the future. Artificial intelligence is a tool for economic growth; however, there are also concerns about the negative impact that could result from its use; for example, in the reduction in the number of employees after the introduction of robotics and artificial intelligence.[25]

The Czech News Agency – A Pioneer of Automated Journalism in the Czech Republic

While major media organisations in English-speaking countries use elements of artificial intelligence more often to create journalistic content, media landscapes of smaller language markets, such as the Czech Republic, are in the early stages of developing automated journalism. We have already mentioned that there may be several reasons for this – from the limited number of news texts of individual thematic domains to which machine learning elements could be applied to the relatively high costs of developing creative software for which there is no money left in Czech news media's budgets. *The Czech News Agency* (ČTK)[26] has pioneered the automatic generation of news texts in the Czech Republic. It was first used in October 2018 to cover the municipal and senate elections.[27]

The Czech Statistical Office (ČSÚ) published election results and at that time computer algorithms created reports based on a pre-set template, in which they placed the published data directly. About half of the headlines of the election result service were robotically prepared, as well as most of the fleets (short reports consisting of two paragraphs with a forecast of the final result) and also reports after counting one hundred percent of the votes.[28] A total of over 200 automatically generated texts passed through the hands of the editors (Picture 1).[29]
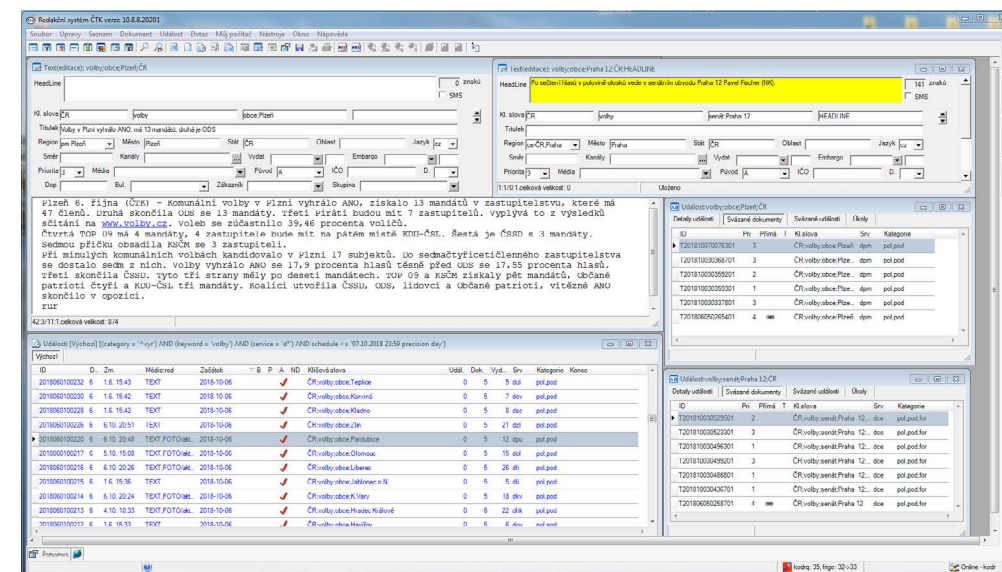
Picture 1: The newsroom programme of The Czech News Agency

Source: Own processing, photographs taken during the participating observation in The Czech News Agency

Preparation for the implementation of automated journalism in *The Czech News Agency* took approximately two months and was tested on data from the previous elections. *"The main benefit of using*

19   MONTAL, T., REICH, Z.: I, Robot. You, Journalist. Who Is the Author? In *Digital Journalism*, 2017, Vol. 5, No. 7, p. 829–849.
20   ANDERSON, C. W.: What Aggregators Do: Towards a Networked Concept of Journalistic Expertise in the Digital Age. In *Journalism*, 2013, Vol. 14, No. 8, p. 1008-1032.
21   CLERWALL, C.: Enter the Robot Journalist: Users' Perceptions of Automated Content. In *Journalism Practice*, 2014, Vol. 8, No. 5, p. 519-531.
22   WEEKS, L.: Media Law and Copyright Implications of Automated Journalism. In *Journal of Intellectual Property & Entertainment Law*, 2014, Vol. 4, No. 1, p. 67-94.
23   BECKETT, C.: *New Powers, New Responsibilities. A Global Survey of Journalism and Artificial Intelligence*. London : London School of Economics and Political Science, Polis, 2019, p. 12.
24   DÖRR, K. N.: Mapping the Field of Algorithmic Journalism. In *Digital Journalism*, 2016, Vol. 4, No. 6, p. 700-722.
25   FREY, C., OSBORNE, M.: *The Future of Employment: How Susceptible Are Jobs to Computerization*. Oxford : Oxford Martin

School, University of Oxford, 2013. [online]. [2019-12-03] Available at: <https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf>.
26   Remark by the authors: *The Czech News Agency* is one of the three public service media in the Czech Republic.
27   ČTK: *ČTK při volbách poprvé využila automaticky generované zprávy*. [online]. [2019-12-04]. Available at: <https://www.ctk.cz/novinky/?id=2764>.
28   MORAVEC, V.: *Proměny novinářské etiky*. Prague : Academia, 2020, p. 85-86.
29   ČTK: *ČTK při volbách poprvé využila automaticky generované zprávy*. [online]. [2019-12-04]. Available at: <https://www.ctk.cz/novinky/?id=2764>.

this new system was to speed up the news service when data was flipping from the Czech Statistical Office, reduce human errors and make it easier for reporters to focus more on the responses and other elections," Radka Matesová Marková, editor-in-chief of ČTK, described in an interview the first experience with automated journalism in the Czech Republic.[30] The editors faced a large number of texts that had to be checked and published as quickly as possible during the vote count.

Processing of texts, from the technological point of view, can be divided into two areas. *The Czech News Agency* used the first area where simple algorithms work on the basis of templates. The process is similar to completing a form, with some conditional elements that can be selected from a finite set of sentences based on data values. The second is artificial intelligence based on machine learning. This may include language processing (structure, grammar, style), natural language processing (NLP), selection and a way of presenting data or understanding texts. The main advantages of using artificial intelligence are speed, error reduction and the extent of processed content.[31] *The Czech News Agency* established the second area in cooperation with researchers in 2019, when it began to address the role of automatically generated news from the Prague Stock Exchange (PSE).

Methodology

We have combined several methods in our research. We chose a combination of questionnaire surveys, interviews, participant observation and experiments (algorithm creation). We prepared a questionnaire survey for ten employees of *The Czech News Agency*, who deal with stock market reporting. We combined the questionnaire survey with structured interviews. In *The Czech News Agency*, we also participated in observation, which was conducted from 5th November to 8th November 2019. After this visit, we were able to describe very precisely journalistic routines and reporters' procedures for creating reports from the Prague Stock Exchange. The research poses six main research questions:

RQ1: What organisational and working routines influence the form of *The Czech News Agency* reports about the results of the trading day on the Prague Stock Exchange?

RQ2: What space for creativity and originality is available for *The Czech News Agency* reporters when reporting about the results of the trading day on the Prague Stock Exchange?

RQ3: Which factors influence the speed of publication of *The Czech News Agency* reports on trade relations on the Prague Stock Exchange?

RQ4: What are the positive and negative aspects of the texts generated by the new software?

RQ5: Can this software speed up the creation of *The Czech News Agency* reports from events on the Prague Stock Exchange?

RQ6: How could we use the time that robotic journalism will save for human journalists in *The Czech News Agency* when creating news from events on the Prague Stock Exchange?

The following paragraphs explain how the experiment was formed.

Rule-based News Generator

The stock exchange news is generated from the obtained current PSE data by a **stochastic context-free grammar-driven generator**. Such generators produce text by constructing a derivation tree from a given context-free grammar and processing it in a top-down mode. The grammar can be predefined and put down by an expert or learned from data using the AI technique called **inductive grammar inference** – an algorithm which makes it possible to learn general grammar rules from a training dataset of suitable examples.

We employed a combined approach. We used the stock exchange news written by human news writers working at *The Czech News Agency*, as the training set for grammar construction. During the first stage, the dataset was thoroughly analysed using natural language processing (NLP) techniques. The n-gram frequency analysis revealed the most typical expressions, phrases and their co-occurrences as used by human journalists. Subsequently, the grammar was compiled by an NLP expert in cooperation with an experienced journalist who assessed the utterances produced by applying the proposed grammar from the point of view of the prospective readers' acceptance.

The generator is implemented as a so-called Mealy machine, i.e. a finite-state automaton (FSA) which emits outputs on transitions from state to state taking into account not only the triggering input but also the state the automaton was in. However, thanks to the extent of the training dataset, it is, at certain transitions, necessary to choose from several different applicable expressions or derivation tree branches (i.e. grammar rules). At this point, a stochastic process is applied.[32]

The stochastic process is used to select specific grammar rules in situations where more than one rule is applicable. The terminal symbols (i.e. the actual words in certain positions) are chosen from sets of suitable expressions and from within the fitness bounds (in the case of indeterminate quantifiers, see below) randomly, too.

As a result of the above-described process, a halting or ponderous message may come out when the randomly chosen words of which the meanings are too similar or are not usually used in the Czech language (in specific collocations or at all). Therefore, a smoothing is performed as the last step of the text generation. The smoothing algorithm simply replaces frequently repeated words.

This approach can potentially encompass every and any procedure which can be formalised as a condition and a subsequent action – in this case, emitting a part of the news. In the journalistic process of creating news, there is often no hard rule for the right decisions and subjective points of view which come up. A reader cannot often reconstruct the original data from the used wording. The semantic charge (or rather its quality) of Czech indeterminate quantifiers like "mírně" (slightly), "nepatrně" (imperceptibly), "výrazně" (massively), "zásadně" (substantially), "významně" (significantly), etc. is naturally fuzzy and thus there is no widely known deterministic rule for how to objectively interpret these expressions. This also leads to problems when algorithmizing the process of creating news.

We need to set some numeric thresholds in order to use the indeterminate quantifiers as mentioned (Figure 2). In a rule-based approach, a human expert can make up and set this number. Nonetheless, there are still problems: What number should it be? How many numeric intervals should cover the variety of quantifiers? These design choices can be challenging and answering these questions is crucial in a rule-based approach.

End-to-End Approach

Generative Adversarial Networks
We also attempted to use neural nets for text generation learned end-to-end. We employed the so-called Generative Adversarial Network (GAN) architecture for generating headlines.[33] However, because of the size

30   MORAVEC, V.: *Proměny novinářské etiky*. Prague : Academia, 2020, p. 86.
31   LEWIS, S., GUZMAN, A., SCHMIDT, T.: Automation, Journalism, and Human-Machine Communication: Rethinking Roles and Relationships of Humans and Machines in News. In *Digital Journalism*, 2019, Vol. 7, No. 4, p. 409-427.

32   MEALY, G.: A Method for Synthesizing Sequential Circuits. In *The Bell System Technical Journal*, 1955, Vol. 34, No. 5, p. 1045-1079.
33   GOODFELLOW, I. et al.: Generative Adversarial Networks. In *Proceedings of the International Conference on Neural Information Processing Systems*. NIPS, 2014, p. 2672-2680. [online]. [2019-12-02]. Available at: <https://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>.

of our dataset and the high variability of phrases used across the news, this experiment did not lead to usable outputs.[34]

Served Model and Connections to Data Sources

The Prague Stock Exchange (PSE) closes down at **16:30** CET every business day. Right after that, we download the available data from the **Prague Stock Exchange API** (Application Programming Interface). It contains the last prices of commodities and **index values** with a timestamp. The **Prague Stock Exchange** does not provide information about the volume of trades in the API, and thus, this specific information is not included in the first draft of the summary of the day published a few seconds after the stock exchange closing.

On top of this, the **Prague Stock Exchange** cannot guarantee the **correctness and finality** of the provided information. Historically, there have been several cases of releasing incorrect numbers after the stock exchange closing reported by the PSE authorities caused by incomplete trades. As a consequence, the PSE decided to release day reports sometime after the stock exchange closing. At this moment, journalists also use the PSE web pages as a source of information for news writing. The **Prague Stock Exchange's** business model introduces a security delay in publishing information on the publicly accessible website. This means that a human writer should wait 15 minutes to be sure that they have reliable material for their work.
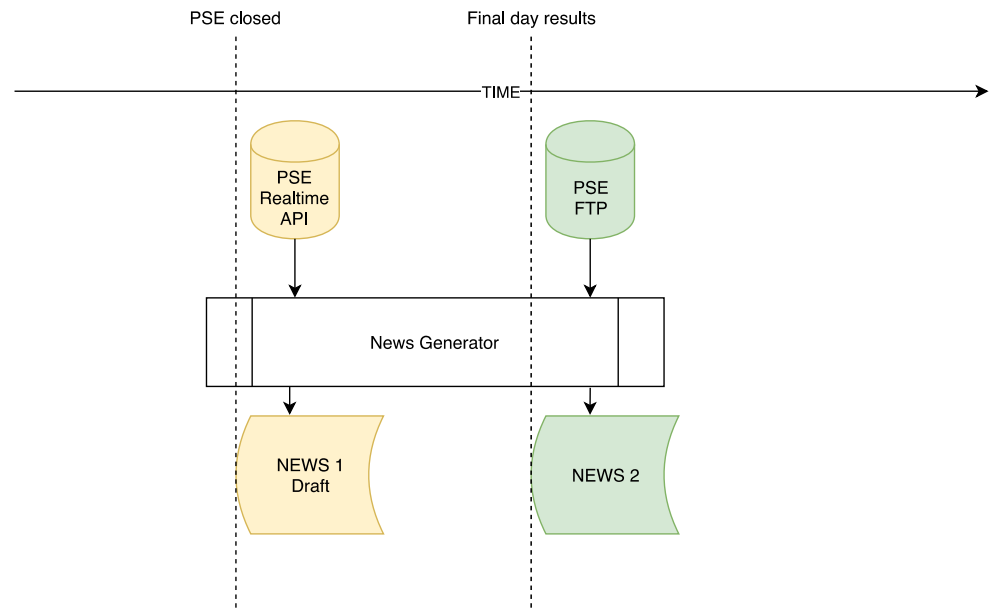

Figure 2: Diagram of the algorithm of automatically generating reports on trading results on the Prague Stock Exchange
Source: Own processing

After the Prague Stock Exchange trading results were generated in Czech by the creative software without human intervention, the next phase of our experiment occurred. The software was implemented at the turn of November and December 2019 into the content newsroom system of *The Czech News Agency* in order to discover how the generated news content was perceived by *The Czech News Agency* exchange newsletters after 14 days of experience.

Automated and robotic journalism in the field of economic and stock reporting has become one of several thematic domains in English-speaking countries. Therefore, in our experiment, we decided on this area in the Czech environment. The birth of the robotic reporter was preceded by a dataset.

We took news released by *The Czech News Agency* during the time period from August 2009 to March 2019 containing 4,677 pieces of news. *The Czech News Agency* routinely releases a summary of the market every day plus special news at the end of every week or on other notable days, e.g. the end of the year. We decided to focus on the summary created as a recap after ordinary business days. The Prague Stock Exchange supported our effort by giving us access to the archive of data files summarising market behaviour.

We prepared a dataset from the archive of *The Czech News Agency* containing 4,677 pieces of news from the Czech stock exchange domain. However, for the purpose of training an algorithm for automatic text generation using structured data, we had to align these pieces of news to reports originating directly from the Prague Stock Exchange archive between January 2012 and October 2019, containing 1,954 data files. As mentioned before, the CNA dataset contains special coverage for certain special events like summaries at the end of the year. Such samples were removed from our dataset as well. The process obviously caused a drop-off. Nonetheless, for the analysis of news structure and domain-specific phrases, all 4,677 pieces of news can be used.

Stock Reporters of The Czech News Agency – Creating Their Reports and Professional Routines

The Czech News Agency's economic service includes a report on trading results on the Prague Stock Exchange every weekday. Reporters of *The Czech News Agency* cover the development of the main of the three indices – the PX Index, which represents the weighted ratio of the thirteen most liquid shares traded on the Prague Stock Exchange and whose calculation also takes into account dividend yields. Reports on trading results on the Prague Stock Exchange represent approximately 3.5% issued by HEADLINE (super-short message with a maximum of 144 characters) of *The Czech News Agency's* economic service and 1% of the classic NEWS REPORT of the same service.

Seven people participate in their creation out of a total of sixteen who create the complete content of *The Czech News Agency's* economic news.[35] They are experienced employees of the news agency, where even the least experienced member has been working for eight years and the most experienced one has been working for *The Czech News Agency* for 21 years.[36] Only one reporter covers stock reporting every day. All seven *Czech News Agency* editors reporting on the results of trading on the Prague Stock Exchange are part of our ethnographic research, which was carried out in November and December 2019 at *The Czech News Agency's* central office in Prague. We posed three main research questions during the first phase:

RQ1: What organisational and working routines influence the form of *The Czech News Agency* reports about the results of the trading day on the Prague Stock Exchange?

RQ2: What space for creativity and originality is available for *The Czech News Agency* reporters when reporting about the results of the trading day on the Prague Stock Exchange?

RQ3: Which factors influence the speed of publication of *The Czech News Agency* reports on trade relations on the Prague Stock Exchange?

The number and form of *The Czech News Agency's* news content on the results of the trading day on the Prague Stock Exchange are based on the commissioning of the agency's editorial board. HEADLINE is

---

34   HU, Z. et al.: Toward Controlled Generation of Text. In *Proceedings of the 34th International Conference on Machine Learning*. PMLR, 2017, p. 1587-1596. [online]. [2020-03-06]. Available at: <http://proceedings.mlr.press/v70/hu17e/hu17e.pdf>.

published first and then the classic NEWS. This happens every weekday as soon as possible after the closing of the exchange (which is at 16:30 CET). The range of the two types of output is different and the content management system technically adjusts the length of HEADLINE, which must not exceed 144 characters, and the CAPTION, which is limited to a maximum of 65 characters. Custom NEWS scope is not modified. Both news genres define the editorial rules of the agency "Principles for Creating The Czech News Agency News".[37] This document will be introduced to new reporters during their training course. The editorial board checks compliance with these principles and then everything is discussed during regular meetings with the staff of the central editorial board. When *The Czech News Agency* editors referred to editorial rules in our research, they often used the phrase "ČTK Bible".[38] The importance of HEADLINE is described in the internal rules for creating ČTK news as follows: *"Its purpose is to significantly speed up the transmission of the first short information to the client and at the same time to create a service that is directly applicable to mobile phones, TV headlines, and other ČTK services or its clients. In verbal reporting, it is a 'super-short report', a concise but comprehensible wording of the main news of the future report. This first news (up to 144 characters) are quick, but necessarily fragmentary."*[39] The maximum number of characters is set for HEADLINE, but the number of sentences is not specified. That is why we encounter various forms of this genre in *The Czech News Agency* stock market news – two sentences, which is usually a set of sentences, or three sentences, where the sentences are simple.

In relation to the NEWS REPORT, ČTK's editorial rules state that, *"there is no universal guide to building a news report"*; however, in *The Czech News Agency*'s textual news it is possible that any subscriber can cut the news, and therefore a well-written report should allow such shortening. The shortening should take place from the end of the news report. In *The Czech News Agency*, this principle is called the "inverted pyramid". *"The meaning of the information corresponds to its location. Less relevant information is placed at the end of the report. Each paragraph can be the last paragraph,"* these editorial rules describe the logic of agency news building.

The introduction of the NEWS REPORT itself must constitute a separate piece of information answering the questions of what, when and where, or who, after separating the rest of the text. In the jargon of the agency this is called a "hard forehead", because it communicates basic information and its source. Our qualitative analysis of *The Czech News Agency*'s 2019 stock market reports confirmed that regardless of the author, the NEWS REPORT has the nature of a template, indicating in the first paragraph the change in the PX index compared to the previous day, which shares had the biggest impact on it, which shares showed the opposite trend of the PX index movement, and the last sentence of the first paragraph of the NEWS REPORT refers to the source, which is the Prague Stock Exchange's website.

Our ethnographic research conducted at *The Czech News Agency* in November and December 2019 showed that the final form of trading reports on the Prague Stock Exchange is influenced by routines that are part of the socialisation process in the newsroom, when senior business reporters transmit values and the standards of their work to younger colleagues.[40] *"When I started here, I learned how to write an article from older colleagues. I know how the stock market report looks from them. Even today I see how other colleagues write news from the stock exchange,"* one of the stock reporters told us.[41] Another reporter said: *"We learn from ourselves. The first paragraph contains basic information on the results of trading and the movement of the PX index, the second paragraph includes the most important changes in share prices, which had a major impact on the development of the PX index, the third notes the share prices with the opposite trend. However, it depends on how the stock market is developing. But there is no precise description of the message structure, it depends on the creativity of each of us."*[42]

Searching for an answer to RQ1 proved that writing stock exchange reports and their resulting form at *The Czech News Agency* has a highly routine form that allows the application of automated or robotic journalism in the Czech environment. This is related to the response to RQ2. The level of creativity and originality of the work on the stock market newsletters is considerably limited. Reporters have the opportunity to use their creativity only where they can influence stock selection, quality and scope of contextual information (e.g. whether to include comments from analysts evaluating the PSE[43] trading day or how long the time series of trend analysis of PX index or individual shares will be). However, it is contextual information in which factual errors may arise, as revealed by our experiment. Their detection is quite complicated even for editors.[44]

Human Fails

During the experiments and the whole process of implementing a usable pipeline, we have found several hotspots in the human approach. Some minor mistakes showed up in trend interpretation (A, B). It could also be understood as manipulating non-systematically with numbers in some cases (B).

Using the end-to-end approach[45] is theoretically possible; however, in narrow domains like the stock exchange, weather forecast or sport reporting, it is too complex a way considering the size of dataset we had.[46] Nonetheless, the rule-based approach is well-suited for these domains provided that we can exactly formalise the decision-making process of human journalists.

Our conclusion is that there is no barrier to using automatic instead of human news generation in mechanical tasks and thus serving the structured data in a more 'human-like' way to the news recipients. There is an important benefit to automated rule-based journalism – errors are systematic and they can be fixed permanently by fixing the procedure. Thanks to this, the same error will never be repeated.

A. **The Czech News Agency reported two days in a row that the Prague Stock Exchange strengthened for the third time.**

190129 T201901290802201 Pražská burza **potřetí** v řadě posílila. Index PX stoupl o 0,44 procenta na 1029,60 bodu. Nahoru burze pomohly hlavně akcie Monety a Komerční banky.

190130 T201901300830001 Pražská burza **potřetí** v tomto týdnu rostla. Index PX dnes posílil o 0,52 pct. na 1034,95 bodu. Dařilo se většině hlavních titulů v čele s Erste Bank.

B. **Two diametrically different reports published on the same day less than an hour apart.**

150731 2015-07-**31 15:21:44** Burza zakončila týden **nepatrným růstem**, PX vzrostl o 0,08 pct na 1031,47 bodu. Z hlavních emisí rostly **ČEZ a KB, Erste a VIG** ale naopak oslabily.

150731 2015-07-**31 16:13:44** Pražská burza zakončila týdenní obchodování téměř **beze změn**, index PX přidal 0,1 pct na 1031,47 bodu. Nejvíc posílily akcie **CME, nejvíc ztratila NWR.**

The response to RQ3 revealed that there are external and internal factors that affect the speed of publication of trading day results on the Prague Stock Exchange. The external factors include the availability

37   Remark by the authors: *The Czech News Agency*'s internal document "Principles for Creating ČTK News". *The Czech News Agency* provided this document for our research.
38   Note: Detected during participant observation (5th November 2019).
39   Remark by the authors: Quoted from *The Czech News Agency*'s internal document "Principles for Creating ČTK News". *The Czech News Agency* provided this document for our research.
40   MCQUAIL, D.: *Žurnalistika a společnost*. Prague : Karolinum, 2016, p. 94.
41   Note: Detected during participant observation (7th November 2019).
42   Note: Detected during participant observation (7th November 2019).

43   Remark by the authors: All stock reporters receive analysts' comments in their e-mail every weekday. According to the answers of our research respondents, this is a limited circle of economists from five major banks and analytical and consulting companies (e.g. Komerční banka, Patria Finance, Next Finance, etc.).
44   Note: Detected during participant observation (7th November 2019).
45   LEWIS, M. et al.: *Deal or No Deal? End-to-End Learning for Negotiation Dialogues*. [online]. [2020-03-06]. Available at: <https://arxiv.org/pdf/1706.05125.pdf>.
46   HU, Z. et al.: Toward Controlled Generation of Text. In *Proceedings of the 34th International Conference on Machine Learning*. PMLR, 2017, p. 1587-1596. [online]. [2020-03-06]. Available at: <http://proceedings.mlr.press/v70/hu17e/hu17e.pdf>.

and quality of data sources from which *The Czech News Agency* exchange reporters draw information. Since 2016 they have drawn from the PSE website. On this site, business data is delayed by 15 minutes due to the PSE's trade policy.[47]

The Czech News Agency exchange reporters thus start writing their reports at 16:45 CET, a quarter of an hour after the PSE closes. Our observation and subsequent quantitative analysis revealed that when the Prague Stock Exchange began to change the look and structure of its website in the last quarter of 2019, this was reflected in the times of HEADLINES and NEWS REPORTS on the results of the trading day.

On some days, the data at the end of the trading day was not accessible. Another external factor is information and comments from analysts. There is a delay in completing a report that spans and rewrites, which is more extensive than when it does not contain analysts' comments. The internal factors include the experience and adoption of routine procedures by stock reporters as well as the workload of the editor who is responsible for reviewing the report and its subsequent publication.

We have decided to verify the results of our observation through quantitative research comparing the times of HEADLINES and NEWS REPORTS on the results of the trading day on the Prague Stock Exchange. We had the release times of 2,366 HEADLINES and 2,285 NEWS REPORTS published by *The Czech News Agency* on its service between 2nd May 2006 and 13th December 2019. The relevant data for our research was only from 2016, when the work of *The Czech News Agency* exchange reporters did not change organisationally and source-wise. Everything is calculated from the reference time of 16:35 CET, when the trading data is final after the closure of the Prague Stock Exchange (see Figure 3).
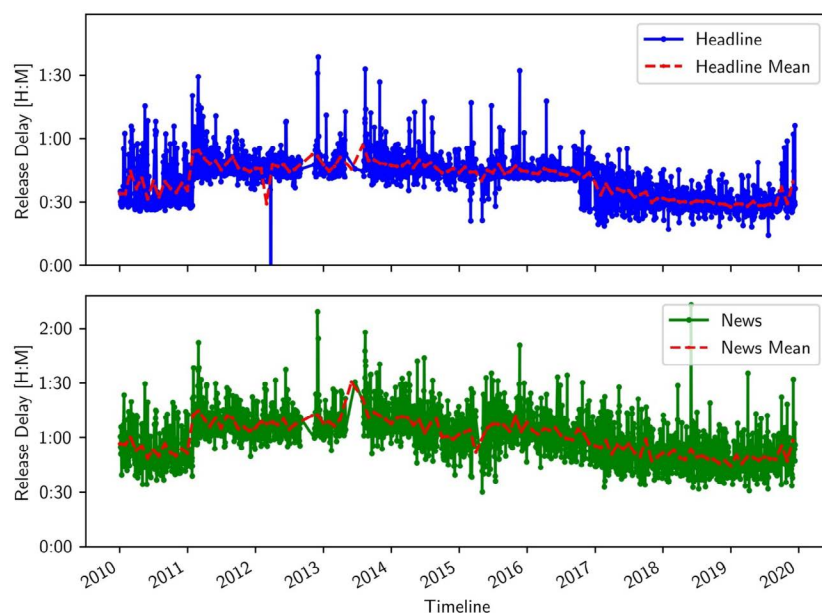

Figure 3: Comparison of the times of origin and publication of HEADLINES and NEWS REPORTS on the results of the trading day on the Prague Stock Exchange by The Czech News Agency
Source: Own processing

Figure 3 shows that the adoption of routine procedures by *The Czech News Agency* exchange reporters led between 2016 and 2019 to a shortening in the time of origin and issue of HEADLINES and NEWS REPORTS by an average of a quarter of an hour. By the end of 2019, several events occurred that increased the

47   *Akcie.* [online]. [2020-03-20]. Available at: <https://www.pse.cz/udaje-o-trhu/akcie/prime-market>.

time of origin and issue of HEADLINES and NEWS REPORTS again. In connection with the modification of the Prague Stock Exchange website, the time to publish increased again (by an average of ten minutes), so that the issue of HEADLINES and NEWS REPORTS was almost comparable to the time before 2016.

Robotic Reporter vs. Human Stock Reporter: Competitors or Partners?

Seven reporters and three members of the management team of *The Czech News Agency* were involved in our qualitative research – the editor-in-chief, the chief editor and the head of the central editorial office. At this stage we tried to answer three research questions:

RQ4: What are the positive and negative aspects of the texts generated by the new software?

RQ5: Can this software speed up the creation of *The Czech News Agency* reports from events on the Prague Stock Exchange?

RQ6: How could we use the time that robotic journalism will save for a human journalist in *The Czech News Agency* when creating news from the events on the Prague Stock Exchange?

All ten respondents agreed in their responses to RQ4 that the strengths of our creative software output are its speed and accuracy. *"Accuracy of defined data, speed,"* said the editor-in-chief. *"After initial embarrassing attempts, they already resemble common reports that could be published after minor edits,"* added the chief editor of *The Czech News Agency*. All respondents identified a weakness in the text output of our software as the weak existence of a comparison of longer time series of the PX index and stock prices, as well as the lack of analyst comments. *"It would be appropriate to work with a number of previous changes to the PX index, and the news would be enriched and diversified. For comparison, yesterday's headline robot: 'Prague Stock Exchange stagnated', and human reporter: 'Prague Stock Exchange has grown again, but today only a tenth of a percent',"* said one of the stock reporters at *The Czech News Agency*.

Another stock reporter said on the weaknesses of the software: *"I miss a longer-term comparison of the development of the issue or the entire stock exchange. Interesting information is if the stock exchange weakens for several days in a row, whether the Index PX reached a record value or if any issue reached the lowest/highest price in a longer period. I also lack the missing context for the development of emission prices. They may weaken or grow, for example, due to the announcement of business results or significant acquisitions."* While describing the weaknesses of our software outputs, ČTK's editor-in-chief said: *"The algorithm does not know a broader generalisation yet (e.g. sectoral), evaluating some phenomenon in a longer time horizon than one day. It also does not evaluate market activity or volumes. It can reliably do most/least, but will not evaluate which title is how important."*

Other reservations were related to the formal aspects of the generated texts. *"The report from the robotic reporter shows at first glance that it was not made by someone with a sense of language. The setting is noticeable. For example, when the stock exchange strengthened, it helped (logically) to grow some shares, when it lost, some stocks pulled it down,"* said a stock reporter who has been working for *The Czech News Agency* for 19 years. *"There are still formal flaws such as the dot at the end of the headline and repetition of words, which is unnecessary,"* added the chief editor.

Three-quarters of respondents agreed in their RQ5 responses that the software could significantly speed up and refine *The Czech News Agency* stock market news – at least 10 to 15 minutes for both HEADLINES and NEWS, compared to today. *"Of course, they'll speed up if there's no need to edit them too much. Writing a stock market report takes about 20 minutes, so if the algorithm were very handy, it could be about a quarter of an hour. This is assuming that it still needs to be checked,"* said one of the stock reporters. *"Yes, it will*

speed it up. Numbers drop into the message itself, compare. Everything can be stylistically adjusted," added another *Czech News Agency* reporter from the stock exchange. The editor-in-chief of *The Czech News Agency* spoke about the application of automated journalism elements to the stock market news: *"If the algorithm is further improved, it could accelerate it, but its current form serves mainly as a reliable basis, which needs to be supplemented with another context. The algorithm also fails to take into account the impact of current events on the market, which is understandable – this is already a journalistic evaluation. We also usually add some analysts' comments to the reports. In my opinion, the savings could be about 10 – 15 minutes in the future, if the algorithm can be improved on the basis of specific comments from colleagues who write news from the Stock Exchange."*

Only stock reporters had sceptical views, not editorial leadership. *"We could take the preliminary message from the robot as the default text, but it would probably take longer to edit it than writing your own text,"* said one of them. Reporters and members of *The Czech News Agency*'s leadership in response to RQ6 saw the promise of relief from routine and less creative journalistic work from the application of automated and robotic journalism to *The Czech News Agency* stock reporting. *"The editor can perform other, more creative tasks,"* said the chief editor. *"He could devote himself to the creation of other texts that require more journalistic input,"* added the editor-in-chief. Similar answers were heard from stock reporters. *"In some cases, time savings would be useful when you need to work quickly on something else,"* said reporter. *"Time can be used for other, more creative work,"* another added. *"The saved time can be used to find additional information that would broaden and humanise the stock market report,"* also said one of *The Czech News Agency* exchange reporters.

Discussion and Conclusion

Journalists are supposed to use interesting phrases to attract readers and not bother them by schematically repeating the same skeletal texts. Nonetheless, that is the main reason why the end-to-end methods failed on this relatively small dataset in Czech. Narrow domains such as the stock exchange, weather or traffic have the potential for building up an automatic system for generating news by analysing previous (man-made) ones, defining grammar and creating automatic generators using structured data.

The debate on automated journalism has focused on algorithms because the financial situation in the Czech newsrooms suggests that the automation of journalism will be necessary for journalism to continue its mission. Our research confirms that journalists expect that their role will remain important, and that they will work with software to produce reports.

Now we will improve the software with longer time series, remove phrases not found in Czech (for example, we will expand our vocabulary) and carry out further research on how much the preparation of stock market reports will shorten after it is put into operation. All respondents detected the weakness in the text output of our software and also the weak existence of a comparison of the longer time series of the PX index and stock prices. There are still formal shortcomings, such as repetition of words. On the other hand, respondents appreciated the savings of time they had made while using this software.

The implemented grammar-driven generator with hard-coded expert knowledge is able to cover the vast majority of common needs for reporting on the stock exchange trading. However, reports on some long-time dependencies are still missing. For example, we can use sentences which inform about exceeding extremes. The problem with using these sentences is that journalists are generally not able to precisely define which extremes are interesting for readers, and instead must rely on their subjective feelings.

The Czech News Agency has not yet put into operation automatically generated text without any modifications made by humans. Editors miss context information in the text generated from the robotic reporter. Nowadays, our team of researchers tries to improve the system and *The Czech News Agency* will incorporate automatically generated text into their newsroom. Our experiment should show that journalists can have confidence in the automatically generated text. Therefore, they will not have to edit the text in the future, so the text will be checked directly by the editor.

The system should make it possible to publish routine news items much more quickly than currently when reporters look at the Prague Stock Exchange website. The generated texts will be loaded into the Digital Media Systems publishing programme and will be subjected to the same editorial processes as news items created from the outset by human reporters. The whole process should be speeded up considerably. Journalists have many subjective views (and creativity is important for their work). Future research will focus on their work with automatically generated text. Some of the numerous questions arising could be how much of the automatically generated text will be used by journalists in their work and how they will integrate the new algorithms into the Digital Media Systems publishing programme.

BIBLIOGRAPHY:

Akcie. [online]. [2020-03-20]. Available at: <https://www.pse.cz/udaje-o-trhu/akcie/prime-market>.
ANDERSON, C. W.: What Aggregators Do: Towards a Networked Concept of Journalistic Expertise in the Digital Age. In *Journalism*, 2013, Vol. 14, No. 8, p. 1008-1032. ISSN 1464-8849.
Annual Report 2016. [online]. [2020-03-07]. Available at: <https://automatedinsights.com>.
BECKETT, C.: *New Powers, New Responsibilities. A Global Survey of Journalism and Artificial Intelligence.* London : London School of Economics and Political Science, Polis, 2019.
BLANKESPOOR, E., DEHAAN, E., ZHU, C.: Capital Market Effects of Media Synthesis and Dissemination: Evidence from Robo-journalism. In *Review of Accounting Studies*, 2018, Vol. 23, No. 1, p. 1-50. ISSN 1573-7136.
BODEN, M.: Artificial Intelligence. In AL-KHALILI, J. (ed.): *What's Next?* London : Profile Books, 2017, p. 118-128.
CASWELL, D.: Structured Journalism and the Semantic Units of News. In *Digital Journalism*, 2019, Vol. 7, No. 8, p. 1134-1156. ISSN 167-082X.
CASWELL, D., DÖRR, K.: Automated Journalism 2.0: Event-Driven Narratives. In *Journalism Practice*, 2018, Vol. 12, No. 4, p. 477-496. ISSN 1751-2794.
CLARKE, A. C.: *2001: Vesmírná odysea.* Prague : Laser, 2008.
CLERWALL, C.: Enter the Robot Journalist: Users' Perceptions of Automated Content. In *Journalism Practice*, 2014, Vol. 8, No. 5, p. 519-531. ISSN 1751-2786.
ČAPEK, K.: *R.U.R.* Prague : Artur, 2004.
ČTK: *ČTK při volbách poprvé využila automaticky generované zprávy.* [online]. [2019-12-04]. Available at: <https://www.ctk.cz/novinky/?id=2764>.
DEEMTER, K., KRAHMER, E., THEUNE, M.: Real Versus Template-Based Natural Language Generation: A False Opposition? In *Computational Linguistics*, 2005, Vol. 31, No. 1, p. 15-24. ISSN 1530-9312.
DÖRR, K. N.: Mapping the Field of Algorithmic Journalism. In *Digital Journalism*, 2016, Vol. 4, No. 6, p. 700-722. ISSN 2167-082X.
FRANKLIN, S.: A Foundational Architecture for Artificial General Intelligence. In GOERTZEL, B., WANG, P. (eds.): *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms.* Amsterdam : IOS Press, 2007, p. 36-54.
FREY, C., OSBORNE, M.: The Future of Employment: How Susceptible Are Jobs to Computerization. Oxford : Oxford Martin School, University of Oxford, 2013. [online]. [2019-12-03]. Available at: <https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf>.
GOODFELLOW, I. et al.: Generative Adversarial Networks. In *Proceedings of the International Conference on Neural Information Processing Systems*. NIPS, 2014, p. 2672-2680. [online]. [2019-12-02]. Available at: <https://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>.
GRAEFE, A.: *Guide to Automated Journalism.* New York : Tow Center for Digital Journalism, Columbia Journalism School, 2016. [online]. [2019-12-02]. Available at: <https://academiccommons.columbia.edu/doi/10.7916/D80G3XDJ>.
HAMILTON, J. T., TURNER, F.: *Accountability Through Algorithm: Developing the Field of Computational Journalism.* [online]. [2020-03-08]. Available at: <https://web.stanford.edu/~fturner/Hamilton%20Turner%20Acc%20by%20Alg%20Final.pdf>.

HU, Z. et al.: Toward Controlled Generation of Text. In *Proceedings of the 34th International Conference on Machine Learning*. PMLR, 2017, p. 1587-1596. [online]. [2020-03-06]. Available at: <http://proceedings.mlr.press/v70/hu17e/hu17e.pdf>.

LEWIS, M. et al.: *Deal or No Deal? End-to-end Learning for Negotiation Dialogues*. [online]. [2020-03-06]. Available at: <https://arxiv.org/pdf/1706.05125.pdf>.

LEWIS, S., GUZMAN, A., SCHMIDT, T.: Automation, Journalism, and Human-Machine Communication: Rethinking Roles and Relationships of Humans and Machines in News. In *Digital Journalism*, 2019, Vol. 7, No. 4, p. 409-427. ISSN 2167-0811.

MARCONI, F., SIEGMAN, A.: *The Future of Augmented Journalism: A Guide for Newsrooms in the Age of Smart Machines*. New York : Associated Press, 2017, p. 4. [online]. [2019-12-02]. Available at: <https://insights.ap.org/uploads/images/the-future-of-augmented-journalism_ap-report.pdf>.

MARR, B.: *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*. London : Wiley, 2016.

MAYER-SCHÖNBERGER, V., CUKIER, K.: *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London : John Murray Publishers, 2013.

MCDONALD, D. D.: Natural Language Generation. In INDURKHYA, N., DAMERAU, F. J. (eds.): *Handbook of Natural Language Processing*. Boca Raton : Chapman and Hall, 2010, p. 121-141.

MCQUAIL, D.: *Žurnalistika a společnost*. Prague : Karolinum, 2016.

MEALY, G.: A Method for Synthesizing Sequential Circuits. In *The Bell System Technical Journal*, 1955, Vol. 34, No. 5, p. 1045-1079. ISSN 00058580.

MONTAL, T., REICH, Z.: I, Robot. You, Journalist. Who Is the Author? In *Digital Journalism*, 2017, Vol. 5, No. 7, p. 829-849. ISSN 2167-0811.

MORAVEC, V.: *Proměny novinářské etiky*. Prague : Academia, 2020.

STRAY, J.: Making Artificial Intelligence Work for Investigative Journalism. In *Digital Journalism*, 2019, Vol. 7, No. 8, p. 1076-1097. ISSN 2167-0811.

WEEKS, L.: Media Law and Copyright Implications of Automated Journalism. In *Journal of Intellectual Property & Entertainment Law*, 2014, Vol. 4, No. 1, p. 67-94. ISSN 2324-6286.