ТВОРЧЕСТВО МОЛОДЫХ ИССЛЕДОВАТЕЛЕЙ CREATIVITY OF YOUNG RESEARCHERS

УДК 070:004.8

DOI 10.17150/2308-6203.2021.10(4).734-744

Зорина Виолетта Александровна

Аспирант

Кафедра массовых коммуникаций, филологический факультет, Российский университет дружбы народов, г. Москва, Российская Федерация, e-mail: zorina-va@rudn.ru

Violetta A. Zorina

PhD Student

Department of Mass Communication, Faculty of Philology, Peoples' Friendship University of Russia (RUDN University), Moscow, Russian Federation, e-mail: zorina-va@rudn.ru

Осиповская Елизавета Андреевна

Кандидат филологических наук, доцент

Кафедра массовых коммуникаций, филологический факультет, Российский университет дружбы народов, г. Москва, Российская Федерация, e-mail: osipovskaya-ea@rudn.ru

Elizaveta A. Osipovskaya

PhD in Philology, Associate Professor

Department of Mass Communication, Faculty of Philology, Peoples' Friendship University of Russia (RUDN University), Moscow, Russian Federation, e-mail: osipovskaya-ea@rudn.ru

Обзор литературы на тему искусственного интеллекта в журналистике: библиометрический анализ научных статей, проиндексированных в международных базах данных WoS и Scopus

Аннотация. Искусственный интеллект (ИИ) — быстро развивающаяся сфера в США и Европе, которая затрагивает практически каждую сферу человеческой жизни: медицину, бизнес, экономику, инженерии и медиаиндустрию. Согласно отчету консалтинговой компании *Gartner*, популярность технологий искусственного интеллекта активно росла с 2015 г. Спрос на технологии ИИ закономерно породил повышенный интерес научного сообщества и, как следствие, увеличение числа публикаций на данную тему. В этом исследовании проведен обзор научных статей на тему использования ИИ в области журналистики. В основе методологии исследования лежит работа, посвященная эволюции и концептуализации data-журналистики. Цель исследования — определить степень научной разработанности темы, связанной с использованием ИИ в сфере медиакоммуникаций. Для этого был применен метод количественного анализа, который позволил отобрать по ключевым словам статьи, индексируемые международными наукометрическими базами данных *Scopus* и *Web of Science*, за период с 2015 по 2021 г.

На тему ИИ в медиа было выявлено 130 статей, после удаления дубликатов корпус выборки составил 70 публикаций. Авторы сосредоточились на статьях, основанных на эмпирических исследованиях, и исключили обзоры, которые лишь сообщали о потенциале искусственного интеллекта и его текущем состоянии в журналистике. В дальнейшем статьи были проанализированы по следующим параметрам: 1) частота публикаций за определенный период времени, 2) аффилиация авторов научных статей для установления географического указания, 3) цитируемость. Авторы анализировали, как зарождался интерес к ИИ в медиаиндустрии и достигал своего пика. Согласно результатам, с 2015 по 2018 г. наблюдался умеренный рост числа публикаций, а в 2019 г. произошел резкий скачок публикационной активности. В то же время наибольшее число статей наблюдалось в 2020 г., что объясняется возросшей ролью технологий искусственного интеллекта в эпоху COVID-19. ИИ сыграл ключевую роль во всех аспектах реагирования на кризис, например, в прогнозировании эволюции вируса, создании эффективных инструментов для сферы здравоохранения. Ожидается, что в 2021 г. количество публикаций будет расти, поскольку пандемия сыграла каталитическую роль в цифровой трансформации и эволюции ИИ. Исследование также показало, что первые три позиции по количеству публикаций в двух наукометрических базах занимают ученые из США, Китая и Англии. Лидерство авторов из США во всех анализируемых аспектах неудивительно, их статьи раскрывают теоретические основы и фундаментальные вопросы искусственного интеллекта в журналистике. Было установлено, что ИИ в новостных агентствах применяют для разных задач, например, для процедуры проверки фактов, поиска информации в больших базах данных, написания коротких статей и заметок, автоматической публикации и расстановки тегов. Поскольку каждое действие выполняется машиной, авторы исследований используют следующие термины для описания ИИ в журналистской практике: «автоматизированная журналистика», «роботизированная журналистика» и «алгоритмическая журналистика». В рамках «роботизированной журналистики» ИИ рассматривается в качестве потенциального репортера и редактора. В «алгоритмической журналистике» нейросети используются для сбора информации, создания историй и распространения их среди граждан. «Автоматизированная журналистика» подразумевает анализ больших баз данных, работу модераторов, выявление фейковых новостей и написание простых по структуре статей. Каждая категория подразумевает использование определенного типа технологий ИИ, при этом, в целом, данные понятия синонимичны друг другу. Анализ содержания статей позволил сформулировать предложения для дальнейших исследований. Например, сегодня наиболее остро стоит вопрос, связанный с этической и правовой стороной использования роботизированных технологий в журналистской практике. Не менее актуальным вопросом является и потенциальная вероятность замены журналистов роботами.

Ключевые слова. Искусственный интеллект, автоматизированная журналистика, алгоритмическая журналистика, роботизированная журналистика.

Информация о статье. Дата поступления 6 августа 2021 г.; дата принятия к печати 8 ноября 2021 г.; дата онлайн-размещения 15 декабря 2021 г.

Literature Review on Artificial Intelligence in Journalism: A Bibliometric Analysis of Publications Indexed in the Web of Science and Scopus

Abstract. This article reviews the past-to-present academic literature on artificial intelligence (AI) in journalism. Over the past years, these technologies have attracted the sufficient attention of researchers from various fields of scientific study producing a large number of publications. We have reviewed academic articles published between 2015 and 2021 to provide understanding of the current state of the research on AI in various research areas including journalism. The corpus was gathered by searching publications in two international databases, Scopus and the Web of Science (WoS). 70 empirical studies were selected on the basis of applying AI to journalism. Each article was categorized according to the type of database, period of time, the country of publication, the field of study and the frequency of citations. The applied method of quantitative research allows tracking the development of research within six years in the field of automated journalism. Finally, we put forward several proposals for further research in this field.

Keywords. Artificial intelligence, automated journalism, algorithmic journalism, robot journalism.

Article info. Received August 6, 2021; accepted November 8, 2021; available online December 15, 2021.

Introduction

Artificial Intelligence (AI) is a rapidly evolving sphere in the United States and Europe. Since 1955 when AI was founded by the professor J. McCarthy at Stanford University. There are seemingly endless ways in which AI is affecting people's lives: medicine, economics, business, engineering, and even the media industry. It has already played a role in the development of COVID-19 vaccines by narrowing the field of promising candidates. Self-driving cars are another crucial application area of AI. We interact with AI technology on a daily basis by using smartphones with digital voice assistance. In August 2020, the research company Gartner released a report [1] on emerging trends that are believed to shape the world in the next five years. According to the report the interest in AI has grown tremendously since 2015. Notably, AI-specific technologies were

the first to emerge in the Hype Cycle, followed by Composite AI, Generative AI, Responsible AI, AI-augmented development, Embedded AI, and AI-augmented design (see Fig. 1).

The demand for AI among technology companies and regular users is proliferating. According to Google Trends, searches for «Artificial Intelligence» were extremely high in demand in 2020 (see Fig. 2). Though Google Trends does not show the absolute search volume data, it presents the relative popularity of a search query. We chose to view a «Worldwide» chart to see how users' searches differ from one country to another, then filtered the results by the «Science» category to exclude data related to the entertainment industry. As a result, the United Arab Emirates (UAE) and China were on top of the query. AI is part of the UAE's economic diversification strategy to transform

Fig. 1. Hype Cycle for Emerging Technologies, 2020. Gartner (August 2020) [2]

the country from an oil-dependent economy to a knowledge-based one. UAE's ambition is to become one of the world top centres of AI. Remarkably, in 2017, the UAE became the first country in the world to appoint a Minister of State for AI. China is a number two leader in AI-empowered businesses; it consistently files AI patents more than any other country. For instance, the number of Chinese AI companies dealing in speech and image recognition applications amounted to 1 189 companies in 2019.

The chart shows that the interest rate in AI is high and steady without any peaks and troughs over the year. Google Trends also reveals AI-related topics and queries that encompass "learning artificial intelligence", "artificial intelligence machine learning", "artificial intelligence course", "postgraduate education",

"master of science", "Coursera", "bachelor of technology". Thus, we can assume that such technologies are in high demand in education [3]. Indeed, these tools are frequently used to track student behaviour, predict their performance, and deliver content in an engaging way. There are AI tutors who provide inclusive education and create personalized learning paths.

At the same time, AI-related queries include searches for fake news. The media often uses AI to discern truth from fiction. The most prominent example is Logically, the UK-based startup using AI to combat the spread of misinformation on the Internet and social platforms and to provide a fact-checking service to detect fake news. The algorithms also check the toxicity of posts and can block obscene publications.

Fig. 2. Search topic «Artificial Intelligence» in Google (Google Trends, 2020)*

* Google Trends data reflects searches people make on Artificial Intelligence (2020). Official site. URL: <https://trends.google.com/trends/explore?q=Artificial%20intelligence>.

Methodology

The paper does not present new research but explores the corpus of AI and journalism works. Thus, its purpose is to analyze academic literature on AI in journalism, in particular foreign authors' publications indexed in Scopus and WoS. We evaluated the field development, and the works made an impact on AI conceptualization. The following research questions have been set to achieve the goal:

RQ1. What is the appropriate term for using AI technologies in the media industry?

RQ2. How has the research literature on AI developed over six years?

RQ3. What are the study areas of AI?

RQ4. Who are the most authoritative researchers in the field of automated journalism?

In the current paper, we developed the definition of data-journalism that was used in the measurement process [4]. We applied the method of quantita-

tive research. In the study, we included only academic publications indexed in the Web of Science and Scopus within the period from 2015 to 2021. In order to review the existing literature, we identified the keywords related to AI in the media industry. Then we selected relevant terms from the keyword sections of the papers and compiled the list of words related to AI and journalism. Next, we carried out a literature search in two scientific databases, obtaining 130 records in total. Thereafter, we identified duplicate papers and cleared the database. It resulted in a corpus of 70 publications. Then we selected the relevant publications and coded the material by the period of time, the field of study and geographical affiliation of researchers.

Results

Finally, we described the phenomenon of AI. Scholars several terms

concerning the field of AI in journalism: “robot journalism”, “automated journalism”, “algorithmic journalism” and “automated news”.

Kunert [5], Moravec et al. [6], Segarra-Saavedra et al. [7], Milosavljevic et al. [8] apply the term “automated journalism” to AI in the media, arguing that this is the most usable tool for journalists. Brlek et al. [9], Diakopoulos et al. [10], Dorr et al. [11], Waddel [12] propose “algorithmic journalism” emphasizing that AI is used to gather information and produce stories. The term “robot journalism” is suggested by Salazar [13], Montal et al. [14], Carlson [15], Kim et al. [16]; they emphasize that robots can be real reporters who write stories and can potentially replace both journalists and editors. This, each definition has its strengths and weaknesses but we tend to use “automated journalism” as it highlights the most relevant features of AI technologies in journalism: analyzing large datasets, moderate reader comments, detect

word patterns that may indicate fake news, scan social media for stories and write simple stories based on datasets.

According to Fig. 3, there was a steady but slow growth of media interest in AI from 2015 to 2018, followed by a sharp increase in 2019. The dramatic rise in 2020 can be explained by the COVID-19 pandemic, when AI has played a key role in every aspect of the crisis response, for instance, predicting the evolution of the virus and creating the effective tools for the healthcare industry. Publications are expected to continue to grow over 2021 since the COVID-19 pandemic has played a catalytic role in enhancing the future of digital transformation and the evolution of AI.

The study shows that 27 % of the papers indexed in the Web of Science were written by USA scholars. The intensification of international cooperation stimulates and supports discoveries. For example, China strives to take the leading position in the AI market. Today, it is the second in the “technol-

Fig. 3. The number of articles on AI in journalism in the WoS and Scopus databases, %

ogy race” and, according to the survey, and accounts for only 17 % of the total WoS publications. Interestingly, London heads the CIMI ranking (according to the report of IESE Business School “Cities in Motion Index” that aims to evaluate the development of the world smart cities), though the papers by UK scholars amount to 10 % of the database. The publications by scholars from other countries are presented in a much smaller proportion. The same pattern was discovered in the Scopus database: the majority of papers come from the universities in the USA (24 %), China (14 %) and the United Kingdom (10 %).

Fig. 4 presents AI research areas. Over 50 % of the articles present research in computer science (1 657 out of 5 324) and engineering (1 347 out of 5 324). This is due to the fact that AI is based on digitalization and automation. On the contrary, the papers on AI in journalism are poorly presented in the corpus (social science — 3 %).

The Scopus database (see Fig. 5) reveals a similar trend: most of the articles refer to computer science (22 %) and engineering (17 %). However, the number of publications on social science is bigger (9 %) since there are more Social Science journals in the Scopus database compared to WoS.

In our analysis of AI in journalism, we focused on empirical studies and neglected surveys describing the potential of artificial intelligence.

Fig. 6 shows a steady growth of the WoS-related publications beginning in 2016. Interestingly, just at that time during the Rio Olympic Games, *The Washington Post* initiated an experiment with automated news (so-called “robot journalism”). The machine analyzed the raw data from the Games, then match them to the relevant phrases in the story template and add the information to create a narrative that can be published across different platforms. It was a successful experiment that showed the magnificent opportunities for AI further

Fig. 4. Research areas on AI in the WoS citation database, %

Fig. 5. Research areas on AI in the Scopus citation database, %

Fig. 6. The number of articles on AI in journalism in the WoS citation database, %

use. The experiment proves good only for data-grounded news like sports and finance. Hence, it might give reasons for publications significant growth between 2016 and 2017.

The country-of-origin factor (see Fig. 7) presents that most research-

ers come from institutions in the USA (33 %) and Spain (16 %). American leadership in the technological race explains well the country's high citation index. Additionally, the articles establish a theoretical framework and discuss the fundamental issues of artificial

Fig. 7. The country of articles' origin in WoS citation databas, %

intelligence in journalism. Thus, other scholars cite US authors.

The next step of the research was to explore the most cited papers on AI in the media. One of them “Algorithmic accountability: journalistic investigation of computational power structures” 190 by Professor Nicholas Diakopoulos, the USA describes the phenomenon of algorithmic accountability reporting as a mechanism for elucidating and articulating the power structures, biases, and influences those computational artefacts exercise in society [17]. Another highly cited article, “Automated Media: An Institutional Theory Perspective on Algorithmic Media Production and Consumption” 114 by professor Philip Napoli offers the perspectives of automated journalism and algorithmic media production [18]. Many authors referred to the study “Clarifying Journalism’s Quantitative Turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting”» by professor Mark Coddington, the USA [19]. The study examines the roles of three quantitative forms: journalism-

computer-assisted reporting, data journalism, and computational journalism in contemporary journalistic practice.

Conclusions

The findings show that the overall number of publications on artificial intelligence significantly increased from 216 to 3 159 academic papers in the Scopus citation database, and from 131 to 2 644 in the Web of Science citation database from 2015 to 2021. There was a tendency to explore the sphere of artificial intelligence in the media. There are various ways to use AI-technologies in journalism: to verify data, for fact-checking procedures, to produce stories and journalistic notes, for publishing and automatically tagging articles. Since all these activities are performed by a machine, the authors of the research prefer to use the general term “automated journalism” when describing such processes. The dramatic increase of publications and cases of how robots affecting journalists’ work show the potential of automated journalism. Moreover, the

analysis revealed that articles cover such issues as ethics [20], threats of legal liability [21], biases [22], and readers' perceptions [23]. Therefore, our future studies will deal with the following questions: Will robots replace journalists? What should we do with the copyright? Can AI technologies become a subject of social and political controversy?

REFERENCES

1. Panetta K. 3 Themes Surface in the 2021 Hype Cycle For Emerging Technologies. Available at: <https://www.gartner.com/smarterwithgartner/3-themes-surface-in-the-2021-hype-cycle-for-emerging-technologies>.
2. Panetta K. 5 Trends Drive the Gartner Hype Cycle for Emerging Technologies, 2020. Available at: <https://www.gartner.com/smarterwithgartner/5-trends-drive-the-gartner-hype-cycle-for-emerging-technologies-2020/>.
3. Grinshkun V., Osipovskaya E. Teaching in the Fourth Industrial Revolution: Transition to Education 4.0. *CEUR Workshop Proceedings*, 2020, vol. 2770, pp. 9–15.
4. Ausserhofer J., Gutounig R., Matiassek S., Goldgruber E., Oppermann M. The Datafication of Data Journalism Scholarship: Focal Points, Methods, and Research Propositions for the Investigation of Data-intensive Newswork. *Journalism*, 2017, vol. 21, iss. 3, DOI: <https://doi.org/10.1177/1464884917700667>.
5. Kunert J. Automation in Sports Reporting: Strategies of Data Providers, Software Providers, and Media Outlets. *Media and Communication*, 2020, vol. 8, iss. 3, pp. 5–15. DOI: 10.17645/mac.v8i3.2996.
6. Moravec V., Macková V., Sido J., Ekštejn K. The Robotic Reporter in the Czech News Agency: Automated Journalism and Augmentation in the Newsroom. *Communication Today*, 2020, vol. 11, iss. 1, pp. 36–52.
7. Segarra-Saavedra J., Cristofol F. J., Martinez-Sala AM. Artificial Intelligence (AI) Applied to Informative Documentation and Journalistic Sports Writing. The Case of BeSoccer. *Doxa Comunicacion*, 2019, no. 29, pp. 275–286.
8. Milosavljevic M., Vobic I. Human Still in the Loop Editors Reconsider the Ideals of Professional Journalism Through Automation. *Digital Journalism*, 2019, vol. 7, iss. 8, pp. 1098–1116.
9. Brlek SS., Trivundza IT. Algorithmisation of a National Press Agency: the case of STA. *Javnost-The Public*, 2019, vol. 26, iss. 1, pp. 62–81.
10. Diakopoulos N., Koliska M. Algorithmic Transparency in the News Media. *Digital Journalism*, 2017, vol. 5, iss. 7, pp. 809–828.
11. Dorr K.N., Hollnbuchner K. Ethical Challenges of Algorithmic Journalism. *Digital Journalism*, 2017, vol. 5, iss. 4, pp. 404–419.
12. Waddel T.F. 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*, 2019, vol. 96, iss. 1, pp. 82–100.
13. Salazar I. Robots and Artificial Intelligence. New Challenges of Journalism. *Doxa Comunicacion*, 2018, no. 27, pp. 295–315.
14. Montal T, Reich, Z. I, Robot. You, Journalist. Who Is the Author? Authorship, bylines and full disclosure in automated journalism. *Digital Journalism*, 2017, vol. 5, iss. 7, pp. 829–849.
15. Carlson M. The Robotic Reporter: Automated Journalism and the Redefinition of Labor, Compositional Forms, and Journalistic Authority. *Digital Journalism*, 2015, vol. 3, iss. 3, pp. 416–431.
16. Kim D., Kim S. Newspaper Companies' Determinants in Adopting Robot Journalism. *Technological Forecasting and Social Change*, 2017, no. 117, pp. 184–195. DOI: 10.1016/j.techfore.2016.12.002.
17. Diakopoulos N. Algorithmic Accountability: Journalistic Investigation of Computational Power Structures. *Digital Journalism*, 2015, vol. 3, iss. 3, pp. 398–415. DOI: 10.1080/21670811.2014.976411.

18. Napoli Philip M. Automated Media: An Institutional Theory Perspective on Algorithmic Media Production and Consumption. *Communication Theory*, 2014, vol. 24, iss. 3, pp. 340–360. DOI: 10.1111/comt.12039.

19. Coddington M. Clarifying Journalism's Quantitative Turn: A Typology for Evaluating Data Journalism, Computational Journalism, and Computer-Assisted Reporting. *Digital Journalism*, 2015, vol. 3, iss. 3, pp. 331–348. DOI: 10.1080/21670811.2014.976400.

20. Ananny M. Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. *Science Technology and Human Values*, 2016, vol. 41, iss. 1, pp. 93–117. DOI: 10.1177/0162243915606523.

21. Lewis S.C., Sanders A.K., Carmody C. Libel by Algorithm? Automated Journalism and the Threat of Legal Liability. *Journalism and Mass Communication Quarterly*, 2019, vol. 96, iss. 1, pp. 60–81. DOI: 10.1177/1077699018755983.

22. Sun W.L., Nasraoui O., Shafto P. Evolution and Impact of Bias in Human and Machine Learning Algorithm Interaction. *PLoS ONE*, 2020, vol. 15, iss. 8. DOI: 10.1371/journal.pone.0235502.

23. Graefe A., Haim M., Haarmann B., Brosius, H.-B. Readers' Perception of Computer-Generated News: Credibility, Expertise, and Readability. *Journalism*, 2018, vol. 19, iss. 5, pp. 595–610. 23. Ostrivnaya E.A. Media Education in the Paradigm of the Creative Economy Using the Model of Student Television. *Kreativnaya ekonomika = Creative Economy*, 2018, vol. 12, no. 8, pp. 1213–1220. (In Russian).

ДЛЯ ЦИТИРОВАНИЯ

Зорина В.А. Обзор литературы на тему искусственного интеллекта в журналистике: библиометрический анализ научных статей, проиндексированных в международных базах данных WoS и Scopus / В.А. Зорина, Е.А. Осиповская. — DOI: 10.17150/2308-6203.2021.10(4).734-744 // Вопросы теории и практики журналистики. — 2021. — Т. 10, № 4. — С. 734–744.

FOR CITATION

Zorina V.A., Osipovskaya E.A. Literature Review on Artificial Intelligence in Journalism: A Bibliometric Analysis of Publications Indexed in the Web of Science and Scopus. *Voprosy teorii i praktiki zhurnalistiki = Theoretical and Practical Issues of Journalism*, 2021, vol. 10, no. 4, pp. 734–744. DOI: 10.17150/2308-6203.2021.10(4).734-744.