

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*: 2023-02-10

Deposited version: Publisher Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Pinto-Martinho, A., Cardoso, G. & Crespo, M. (2022). AI and journalism, robot journalism and algorithms. In Tina Bettels-Schwabbauer, Nadia Leihs, Gyula Maksa, Dominik Speck, Annamária Torbó (Ed.), New skills for journalists: Comparative perspectives from Europe. (pp. 157-169). Pécs: Transylvanian Museum Society.

Further information on publisher's website:

https://eda.eme.ro/bitstream/handle/10598/33083/Letoltes14-Journalism.pdf?sequence=1&isAllowed=y

Publisher's copyright statement:

This is the peer reviewed version of the following article: Pinto-Martinho, A., Cardoso, G. & Crespo, M. (2022). AI and journalism, robot journalism and algorithms. In Tina Bettels-Schwabbauer, Nadia Leihs, Gyula Maksa, Dominik Speck, Annamária Torbó (Ed.), New skills for journalists: Comparative perspectives from Europe. (pp. 157-169). Pécs: Transylvanian Museum Society.. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0 The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

TINA BETTELS-SCHWABBAUER, NADIA LEIHS, GYULA MAKSA, DOMINIK SPECK, ANNAMÁRIA TORBÓ (eds.)

NEW SKILLS FOR JOURNALISTS

COMPARATIVE PERSPECTIVES FROM EUROPE

Erdélyi Múzeum-Egyesület • 2022

© Erdélyi Múzeum-Egyesület, 2022

Published by the Transylvanian Museum Society, 2022 Editor-in-Chief: Annamária Biró

Language Editors: Marcus Denton and Derettens OÜ János Nagy (University of Pécs)

Series Design and Technical Editing: Tipotéka Labs It was produced in the F&F Printing House of Gyergyószentmiklós Manager in Charge: Enikő Ambrus

Descrierea CIP a Bibliotecii Naționale a României New Skills for Journalists : Comparative Perspectives from Europe / ed.: Tina Bettels-Schwabbauer, Nadia Leihs, Gyula Maksa,.... - Cluj-Napoca : Erdélyi Múzeum-Egyesület, 2022 Conține bibliografie ISBN 978-606-739-216-6 I. Bettels-Schwabbauer, Tina (ed.) II. Leihs, Nadia (ed.) III. Maksa, Gyula (ed.) 070 654

The publication of this volume was supported by the EFOP-3.4.3-16-2016-00005 project titled "Modern university in a modern city: value-orientation, openness and inclusive view in a 21st century education institution"



Contents

Preface			•			•	. 9
Part 1: Challenges and Opportunities of Digital Journalism							13
Introduction							15
Data Journalism							17
Raluca Radu, Antonia Matei, Anamaria Nicola, Marian Popovici, Mar Emilia Şercan, Oscar Stănciulescu	ıue	la P	reo	tea	sa,	•	
Defining data journalism							17
Data journalism in the participating countries							21
Research results from the curricula analysis and the interview	s.						23
Data journalism in journalism education							23
Data journalism in the media						•	27
Collaborative Journalism							31
Tina Bettels-Schwabbauer, Nadia Leihs							
Defining collaborative journalism	·	•	·	·	•	·	31
Collaborative journalism in the participating countries	•		•		•	•	33
Research results from the curricula analysis and the interview	s.						36
Collaborative journalism in journalism education							36
Collaborative journalism in the media							39

Contents

Innovative Business Models				43
Ana Pinto-Martinho, Miguel Crespo, Gustavo Cardoso				
Defining innovative business models of journalism				43
Innovative business models in the participating countries				47
Research results from the curricula analysis and the interviews				50
Innovative business models in journalism education				50
Innovative business models in the media	•			53
Ethical Challenges				57
Gábor Polyák, Annamária Torbó, Éva Varga				
Defining ethical challenges of journalism				57
Ethical challenges in the participating countries				61
Research results from the curricula analysis and the interviews				63
Media ethics in journalism education				63
Ethical challenges in the media	•			66
Lessons learned – Insights gained				69
Part 2: Journalism Under New Social Conditions				73
Introduction				75
Storytelling in Social Media				79
Interview results				87

	Contents
Graphic Journalism and Comics Reportage	97
Gyula Maksa	
Interview results	104
Improving Democratic Sensibility	111
Klára Smejkal, Lenka Waschková Císařová	
Interview results	117
Foreign Coverage	125
Dominik Speck	
Interview results	132
Covering Migration	141
Susanne Fengler, Dominik Speck, Marcus Kreutler, Monika Lengauer,	
Anna-Carina Zappe	
Interview Results	148
AI and Journalism, Robot Journalism and Algorithms	157
Ana Pinto-Martinho, Gustavo Cardoso, Miguel Crespo	
Interview results	163
Journalism for Voice-Activated Assistants and Devices	171
Miguel Crespo, Ana Pinto-Martinho, Gustavo Cardoso, Wanessa Andrade	
Interview results	177
Verifying and Analysing Fake News	183
Antonia Matei, Mihaela Păun, Marian Popovici	
Interview results	189

7

8	Contents
	Debunking Disinformation
	Interview results
	Lessons learned – Insights gained
	References
	Annexes

AI and Journalism, Robot Journalism and Algorithms

Ana Pinto-Martinho, Gustavo Cardoso, Miguel Crespo (ISCTE – University Institute of Lisbon)

Automated journalism is also known as algorithmic journalism or robot journalism (Dörr 2016; Montal – Reich 2016; Graefe 2016) and consists of news articles generated by computer programmes. Through artificial intelligence (AI) software, stories are produced automatically by computers rather than human reporters. These programmes interpret, organise, and present data in human-readable ways. The process involves an algorithm that scans large amounts of data, selects from an assortment of pre-programmed article structures, orders key points, and inserts details such as names, places, amounts, rankings, statistics, and other figures. The output can also be customized to fit a certain voice, tone, or style (Dörr 2016; Montal – Reich 2016).

Until now, despite it being a growing trend, not that many media outlets worldwide have used automated journalism on a large scale³¹. Pioneer adopters include *The Associated Press, Forbes, ProPublica,* and *The Los Angeles Times.* Early implementations were mainly used for stories based on statistics and numerical figures. Common topics include sports recaps³², weather, financial reports, real estate analysis, and earnings reviews (Montal – Reich 2016). The *Associated Press* began using automation to cover 10,000 minor baseball leagues games annually.³³ Other than sports, the *Associated Press*

³¹ See a range of case studies: https://www.lse.ac.uk/media-and-communications/polis/JournalismAI/Casestudies

³² See an example from *The New York Times* at https://www.techslang.com/how-is-automated-journalismimpacting-the-media/

³³ https://www.poynter.org/tech-tools/2016/the-associated-press-will-use-automated-writing-to-cover-theminor-leagues/

uses automation to produce stories on corporate earnings. In 2006, *Reuters* announced their switch to automation to generate financial news stories on its online news platform (van Dalen 2012). More famously, an algorithm called Quakebot published a story about a 2014 California earthquake on *The Los Angeles Times* website within three minutes after the shaking had stopped (Carlson 2015; Cohen 2015).

Automated journalism is sometimes seen as an opportunity to free journalists from routine reporting, providing them with more time for complex tasks. It also might allow efficiency and cost-cutting, alleviating the financial burden that many news organisations face. However, automated journalism is also perceived as a threat to the authorship and quality of news and a threat to the livelihoods of human journalists. Graefe and Bohlken (2020) try to give some insights on the perception of automated news in the eyes of users in terms of credibility, quality, and readability in comparison to human-written news. In general, their study showed no difference in readers' perceptions of credibility, a small advantage for human-written news in terms of quality, and a huge advantage for human-written news with respect to readability (see also Jung et al. 2017).

By automating routine stories and tasks, journalists are promised more time for complex jobs such as investigative reporting and in-depth analysis of events (Dörr 2016; Montal – Reich 2016). *The Associated Press* stated³⁴ that, through automation, the news agency freed up 20 percent of reporters' time to focus on higher-impact projects. Automated journalism is also cheaper because more content can be produced within less time.

The main criticisms are related to authorship (who should be credited as the author?), credibility (are algorithms "fair and accurate, free from subjectivity, error, or attempted influence", Gillespie 2014), quality (can machines replace human capabilities

³⁴ https://www.adexchanger.com/publishers/associated-press-uses-ai-boost-contentvideo/volume/021 /02/11/powering-ahead-journalismais-mission-in-2021/

such as creativity, humour, and critical thinking?) and, looking to other activities where humans were replaced by real or virtual machines, the topic of employment. But the introduction of AI in journalism is growing, and projects such as JournalismAI intend to contribute to a wider and better use of it, allowing media organisations to explore how they could use AI technologies to approach a series of challenges, and support a growing network of almost 3,000 journalists across the world (Beckett 2021).

Artificial intelligence

Trying to achieve a definition of artificial intelligence, Mueller and Massaron's (2018) approach is a good starting point:

"When thinking about AI, notice an interplay between goal seeking, data processing used to achieve that goal, and data acquisition used to better understand the goal. AI relies on algorithms to achieve a result that may or may not have anything to do with human goals or methods of achieving those goals."

(Mueller - Massaron 2018: 12)

Skiena explains algorithms as "a procedure that takes any of the possible input instances and transforms it to the desired output" (Skiena 2008: 3). Algorithms are a sequence of steps used to solve a problem and they are also as Mueller and Massaron (2017) argue finite, well-defined, and effective. Behind the ability to learn and interpret speech is machine learning technology (Mueller – Massaron 2016). Machine learning "addresses the question of how to build computers that improve automatically through experience" (Jordan – Mitchell 2015: 255). artificial intelligence includes machine learning, but machine learning does not fully define AI.

The true birth of AI as we know it today began with Alan Turing's (1950) publication of *Computing Machinery and Intelligence*, which explored the idea of how to determine whether machines can think. This paper pointed to the Imitation Game involving three players. Player A is a computer and Player B is a human. Each must convince Player C (a human who cannot see either Player A or Player B) that they are human. If Player C cannot determine who is, and who is not, human on a consistent basis, the computer wins. The same problem can be applied to using AI in journalism. For readers, listeners, and viewers AI must be able to develop a news story that cannot be identified as created by a non-human. But, in ethical terms, the discussion on whether media outlets should identify news stories created by AI is on-going, and there is no clear direction on the path to follow. Van Drunen et al. (2019) provide a framework of transparency instruments in the context of the news personalization algorithms.

A continuing problem with AI is too much optimism (Matteson 2015). The problem that scientists are trying to solve with AI is incredibly complex, even if machines were doing all sorts of amazing things, and AI has its greatest success in areas such as logistics, data mining, and medical diagnosis³⁵.

The adoption of artificial intelligence in newsrooms³⁶, either in reporting, production, or distribution of content, requires training, resources, and ethical debate. Training journalists and editors in general concepts related to artificial intelligence and subsequently in specific technical skills is crucial to promote an organisational culture open to the use of this technology. Implementing AI-based solutions also requires the development of a strategic vision, economic investment, interdisciplinary team building and the search for alliances with educational and technological organisations.

³⁵ See also these examples: Why a Leading AI Expert Is So Optimistic About Humanity's Future: https:// asiasociety.org/blog/asia/why-leading-ai-expert-so-optimistic-about-humanitys-future, Ma vs Musk on AI: The Optimistic Versus the Dystopian Viewpoint: https://www.aitrends.com/ethics-and-social-issues/ ma-vs-musk-on-ai-the-optimistic-versus-the-dystopian-viewpoint/

^{36 4} Examples of AI's Rise in Journalism (And What it Means for Journalists): http://mediashift.org/2016/ 09/4-examples-ais-rise-journalism-means-journalists/

The processes developed with AI should be auditable, adjustable, transparent, and traceable, and respond to ethical standards of journalism; the latter is perhaps the most critical aspect of intelligence implementation in the media industry (see also Dörr –Hollnbuchner 2017).

Broussard et al. (2019) focus on the implications of AI for journalism in the larger context of the digitization of media and public life — a transition to apps, algorithms, social media, etc. In that sense, journalists can begin to learn what AI actually is (and is not), and explain such technologies to the public. The same path is designed by Diakopoulos (2019), as he argues that AI is a new medium through which journalists can express and exercise their ethical and normative values through the code they implement. According to Mueller and Massaron (2016), a true AI will eventually occur when computers can finally emulate the clever combination used by nature: (1) genetics: slow learning from one generation to the next; (2) teaching: rapid learning from organised sources, and (3) exploration: spontaneous learning through media and interactions with others.

Algorithms

All algorithms find solutions, the speedier and easier, the better. Using computers to solve problems by employing the appropriate algorithm speeds up the task significantly, which is the reason that the development of new algorithms has progressed so fast since the appearance of powerful computer systems. Algorithms determine how to interpret big data: process input data and create predictable outputs based on the data patterns. The data itself is not predictable. The reason you need AI and machine learning is to be able to see the patterns in the data and make sense of them (Mueller – Massaron 2016). Since the widespread adoption of the Internet, encounters with algorithmic procedures for "information retrieval" –

the activity of getting some piece of information out of a collection or repository of some kind – have become everyday experiences for most people in a large proportion of the world (Rieder 2020).

A simple definition of an algorithm is a systematic set of operations to perform on a given data set — essentially a procedure. The four basic data operations are create, read, update, and delete (CRUD). This set of operations may be the basis of everything you do with a computer. As the dataset becomes larger, the computer can use the algorithms found in an application to perform more work. The use of immense datasets, known as big data, enables a computer to perform work based on pattern recognition in a nondeterministic manner. In short, to create a computer setup that can learn, you need a dataset large enough for the algorithms to manage in a manner that allows for pattern recognition (Mueller – Massaron 2017: 23).

By combining big data with statistics, you can create a machine learning environment in which the machine considers the probability of any given event (but statistics is not the only machine learning method). One aspect that defined big data as "big" is the notion that while a human can learn something from "big data", the magnitude of the dataset makes human recognition of the patterns impossible (or would take a long time to accomplish). Machine learning helps humans make sense and use of big data. Everything in machine learning revolves around algorithms.

So, according to Mueller and Massaron (2017), it is a sequence of steps used to solve a problem. The sequence presents a unique method of addressing an issue by providing a particular solution. An algorithm does not need to represent mathematical or logical concepts, but people most commonly use algorithms in this manner. Some special formulas are also algorithms, such as the quadratic formula. For a process to represent an algorithm, it must be: (1) finite – the algorithm must eventually solve the problem; (2) well-defined – the series of steps must be both precise and understandable; (3) effective – an algorithm must solve all cases of the problem, for which someone defined it. An algorithm should always solve the problem it has to solve.

Robot journalism

Robot journalism depends on AI and algorithms and partly, in machine learning, although this is only part of what a system requires to become an AI.³⁷ The machine learning portion of the system enables an AI to perform the following kinds of tasks: (1) adapt to new circumstances that the original developer did not envision; (2) detect patterns in all sorts of data sources; and (3) create new behaviours based on the recognized patterns, in other words, make decisions based on the success or failure of these behaviours.

In journalism, the robot is not some physical object or device, but a piece of software able to gather data, analyse it, identify the relevant events to build a news story³⁸ (the more unusual or unexpected, so it becomes news with interest for the audience), organise a sequence and then build a narrative using journalism "rules" and best practices. The result must be as good as a story written by a human, and for the consumer, it should not be identifiable as created by a robot. But, for now, fully automated journalism, (e.g. texts or visuals produced by AI alone) is still rather rare, while journalists increasingly rely on AI to facilitate their work (e.g. tools for data journalism).

INTERVIEW RESULTS

Automation processes are increasingly present in our everyday lives. These processes are now transversal to professions and practices, and journalism is not any different. However, there are substantial differences with regard to its use and

³⁷ See examples from *The New York Times, Reuters*, and other media at https://emerj.com/ai-sector-overviews/ automated-journalism-applications/

³⁸ The Rise of the Robot Reporter, The New York Times, https://www.nytimes.com/2019/02/05/business/ media/artificial-intelligence-journalism-robots.html

evolution in journalism in the countries of the study, as outlined by the interviews we conducted.

We interviewed journalists from five countries about AI and journalism, robot journalism and consequently algorithms: The Czech Republic, Germany, Hungary, Portugal, and Romania. Some work for legacy media, like the broadcaster *Bayerischer Rundfunk (BR)* from Germany and the Česká tisková kancelář (the Czech News Agency – ČTK). The other two newsrooms, 24.hu in Hungary and start-up.ro in Romania, are online media and *Público*, from Portugal has both paper and online editions. Despite the differences in each country's evolution of AI and journalism, most of our interviewees agree that the entrance of such techniques into newsrooms will be inevitable, so skills and literacy in these areas are crucial.

AI in newsrooms and media companies

Of the five interviewees, those from the Czech Republic, Germany and Portugal report that AI was being used in their newsrooms. This is not the case in the newsrooms our interviewees work in Hungary and Romania. Only one of the interviewees is dedicated more specifically to work with automation, Steffen Kühne. His media outlet, German regional public broadcaster *BR*, has a specific unit for developing and working with AI produced content:

"The AI + Automation Lab at Bayerischer Rundfunk was founded in March 2020. Our goal is to develop products to improve the news reporting, production, and distribution at our organisation. This includes, but is not limited to, the use of AI-based approaches to automation and publishing."

Steffen Kühne, Tech lead for the AI + Automation Lab at Bayerischer Rundfunk

Kühne also highlights that there are other teams at *Bayerischer Rundfunk* working with AI, especially in the areas of audience development, media production, archives, and platform development: "We are working together with those teams, for instance when it comes to metadata specifications, storage, and access. These are common topics that are crucial for the future development of automated products."

Radka Matesová Marková from the Czech Republic and Rui Barros from Portugal mention that AI is important in their newsrooms, and there is work being done within the AI scope, although neither of their organisations have a specific unit for this purpose. The news stories produced by AI come from cooperation between the newsroom and the IT department.

"The development is carried out by the IT development team, in cooperation with editorial staff as needed. Strategic coordination is between the IT development director and the Editor-in-Chief. The specific algorithms are prepared by the IT development team together with some experienced news editors."

Radka Matesová Marková, Editor-in-Chief, Česká tisková kancelář

Zsuzsanna Dömös points out that the use of AI in journalism and robot journalism is still not a common practice in Hungarian newsrooms:

"Here AI is not yet integrated into newsrooms. Even large media companies do not have adequate resources and access for this kind of solution. Journalism in Hungary is not keeping pace with the evolution of new technologies."

Zsuzsanna Dömös, Technology journalist, 24.hu

According to our interviewees, automated reporting, in the broader sense, can be used to cover a lot of stories that happen regularly and where results can be quantified (be transformed into structured data). The most common examples quoted were reporting about sports, economy, health, weather, and traffic information. The interviewees are unanimous about the potential in using automated reporting but some of them point out that some stories would largely benefit from human editing and context work.

"Often, this type of reporting can still benefit from the work of reporters and editors, providing additional context on why unemployment has dropped or why a basketball game was lost."

Steffen Kühne, Tech lead for the AI + Automation Lab at Bayerischer Rundfunk

The potential of AI and algorithms was mentioned as something that could benefit all kinds of journalistic products. Steffen Kühne says that investigative reports can use AI to make sense of huge document leaks or satellite data, mentioning that video editors can use specialized software to help organise hours of video footage or streamline colour grading. He also points out that there is a large chance for assistive technologies to reduce the number of tedious tasks for media workers.

In the three newsrooms that reported to have been working with AI, automation has been used. One example is the Czech case where the coverage of local and senate election results in 2018 and regional elections in 2020 was made using automation tools. "Since 2020 we are also using automation tools for regular news reports on petrol prices in the Czech Republic and monthly traffic accident statistics", Radka Matesová Marková explains. In the Portuguese case, Rui Barros says that in the last year there was a lot of work being done regarding the reporting of the pandemic situation, and he suggests that automation has been very important:

"I dare say that the COVID-19 pandemic is the world's first major data driven event. Numbers, but also their constant updating, have never been so important to people. And here, the use of these techniques proved to be advantageous because they allowed the automation of processes."

Rui Barros, Data journalist, Público

While stressing the advantages, our interviewees also reported some challenges linked to AI. Some mentioned that AI depends heavily on data quality, which is often problematic along with data availability. The integration of new software into the existing legacy systems can also be an issue, as Steffen Kühne points out:

"Incompatible interfaces or data models are oftentimes responsible for large development and management overhead. Training and deploying AI models in a cloud infrastructure can also be quite expensive and requires in-depth knowledge on how to optimize algorithms for speed, efficiency, and stability."

Steffen Kühne, Tech lead for the AI + Automation Lab at Bayerischer Rundfunk

Another issue noted were problems raised by linguistic complexity since most models are developed for English:

"The worst thing that can happen to a reader is to feel that the text was produced by an algorithm due to some grammatical inconsistency or some edge case that was not thought of when the product was developed. In this aspect, the Portuguese language offers a lot of resistance."

Labour cost is also an issue, because hiring machine learning experts with hands-on experience is no easy task due to high demand and labour costs, highlights Kühne. A solution for this problem, in his experience, at least partially, is networking, for example by participating in meetups or hackathons, or by establishing

Rui Barros, Data journalist, Público

collaborations with universities. This brings us to a very important point that is forcing media companies and universities to collaborate and push for interdisciplinary partnerships.

Education and training in the AI fields

In a field that is still in an embryonic stage in a lot of countries and newsrooms, it is important to try to understand if the training and teaching offered in universities and other training institutions do exist and whether they are adequate. Steffen Kühne, from Germany, and Radka Matesová Marková, from the Czech Republic, report that, to their knowledge, some universities recently started programmes that include automation in their curricula, where Rui Barros from Portugal, says that it may be in certain curricula, but not in a practical perspective (hands-on). Both Barros and Kühne emphasise that it is very important to work on awareness.

"Some colleges, universities and journalism schools have started to include news automation in their curricula. But it would be presumptuous to say that AI and automation have an overly high priority, especially when considering that traditional journalism training, with an emphasis on basic education, ethics and craftsmanship, is still very relevant today."

Steffen Kühne, Tech lead for the AI + Automation Lab at Bayerischer Rundfunk

Nevertheless, some interviewees say that although it is a benefit for journalists to get more programming and AI related skills, they do not have to be experts in this area. Rather, collaboration between experts in the field of journalism, machine learning, software development, product and design is the secret to success, as Steffen Kühne clearly points out. At present, as there is a lack of specific courses in this field, most of the journalists working in this area are self-taught, relying on their hands-on experience, with some of them going further in their education by taking some courses on programming, for example. Radka Matesová Marková states that she has no formal education in the field and gained her experience through self-study, participation in international conferences and participation in the development of automated journalism tools. Besides, she is a member of a team involving three Czech universities working on a three-year project exploring the ethical aspects of robotic journalism. Steffen Kühne also emphasises that most skills required to solve automation problems come from experience. He holds a degree in journalism and did a couple of semesters in computer science. Rui Barros' interest in this area came from his work as a data journalist. Inspired by what he saw in other international news outlets, Barros searched for information on how it was being done and started to "experiment", a key word for all our interviewees.

In general, the importance and inevitability of AI in the newsrooms is recognized by all our interviewees, whether they use such techniques in their own work or not. Also, the interviewees mention the importance of not only being clear about how the automated work is done but also being transparent about the algorithms in use. Most of our interviewees are self-taught about these issues, sometimes reading and looking for information that helps them experiment, so it is clear that this an area where there is a need for training. Yet, training programmes should consist of several levels – from basic to more advanced – also because the implementation of AI into newsroom routines differs considerably between the countries under study.