The application of artificial intelligence to journalism: an analysis of academic production

Sonia Parratt-Fernández; **Javier Mayoral-Sánchez**; **Montse Mera-Fernández**

Nota: Este artículo se puede leer en español en:
http://www.profesionaldelainformacion.com/contenidos/2021/may/parrat-mayoral-mera_es.pdf

How to cite this article:

Parratt-Fernández, Sonia; **Mayoral-Sánchez, Javier**; **Mera-Fernández, Montse** (2021). "The application of artificial intelligence to journalism: an analysis of academic production". *Profesional de la información*, v. 30, n. 3, e300317.

https://doi.org/10.3145/epi.2021.may.17

Article received on 01-03-2021
Final acceptance: 31-05-2021

Sonia Parratt-Fernández
https://orcid.org/0000-0001-8501-3115

Universidad Complutense de Madrid
Departamento de Periodismo y Comunicación Global
Avda. Complutense, 3
28040 Madrid, Spain
sfparratt@ccinf.ucm.es

Javier Mayoral-Sánchez
https://orcid.org/0000-0002-7371-1925

Universidad Complutense de Madrid
Departamento de Periodismo y Comunicación Global
Avda. Complutense, 3
28040 Madrid, Spain
javier.mayoral@ucm.es

Montse Mera-Fernández ✉
https://orcid.org/0000-0003-4702-9786

Universidad Complutense de Madrid
Departamento de Periodismo y Comunicación Global
Avda. Complutense, 3
28040 Madrid, Spain
mmera@ucm.es

Abstract

Journalism has been able to adapt quickly to technological innovation, especially in recent years. The application of algorithms and artificial intelligence (AI) to this discipline is a phenomenon that has developed rapidly in a very short time. This is therefore a research area that, in spite of its short life, deserves special interest. The objective of this review article is to map and analyze the global scientific production on this topic and to identify which countries are most focused on this issue, which areas are studied most and using which methodological approaches, how and where it is evolving, and the gaps present in this research. The review of 358 texts confirms the considerable attention from academia during the last decade, especially between 2015 and 2020, and that the USA is, by far, the country with most publications on this subject. Most of the published works are research articles carried out, above all, using qualitative methodologies. The areas that have attracted the most interest to date are data journalism, robot writing, and news verification. As is to be expected in a developing discipline, others such as the review of the role of the journalist, the personalization of content, or the incorporation of AI into teaching of journalism have not yet been sufficiently explored but surely will be in the near future.

Keywords

Artificial intelligence; AI; Journalism; Innovation; Review article; Media; Algorithms; New information technologies; NIT; ICT; Theory; Methodologies; Data journalism; Robot journalism; Fact-checking.

1. Introduction

The term "Artificial intelligence", or AI, has been conceptualised from different perspectives, but always emphasizing that it is

> "the capability of a machine to imitate intelligent human behavior" (**Aghion**; **Jones**; **Jones**, 2019)

and that it can be applied to many fields, including journalism. Its development –particularly in recent years– is transforming journalism, while at the same time provoking debate around the benefits and risks it can bring, both to the sector and to the journalists who work within it. Not only this, but AI is also bringing new challenges to a field of academic research which, although nascent, is currently producing a range of increasingly interesting lines of investigation.

According to **Cox** (2000), journalists have been using computers to produce news since 1952, when an early model was used to predict the outcome of the US presidential election. Others place the birth of this practice, often referred to as computer-assisted reporting (CAR), in the US newsrooms of the late 1960s (**Carpien**, 2012; **McGregor**, 2013). Journalist Philip Meyer profiled the protesters of the 1967 Detroit riots with the help of a computer-assisted survey. The resulting article won a Pulitzer Prize and is considered the first attempt by a journalist to employ the analytical methods of social sciences, and use a computer as a tool (**Léchenet**, 2014, p. 5). This is how Meyer began the journey into what became known as precision journalism, from which CAR would inherit some tools –especially in the field of investigative journalism (**Coddington**, 2015).

CAR progressed until it was no longer an appropriate term to refer to the diversity of practices into which it had evolved (**Coddington**, 2015, p. 332), and within which lies data journalism (**Bounegru**, 2012). This concept, which also has its roots in graphic design and investigative journalism, began to be used as a replacement for CAR (**Royal**; **Blasingame**, 2016) and became established in the late 2000s (**De-Lima-Santos**; **Schapals**; **Bruns**, 2020) thanks to the falling price of computers, the use of the internet as a source of information and a medium for publishing, and the open data movement (**Léchenet**, 2014, p. 6). Today the term data journalism can be understood both as a process of analysing large databases in order to present news stories with relevant data, and as a product that allows the user, among other things, to engage with news stories by means of visual representations (**Ausserhofer** *et al.*, 2017, p. 4; **Weber**; **Engebretsen**; **Kennedy**, 2018).

Although some of the practices that CAR led to were not new –automated news personalisation dates back at least to the 1980s– it was not until the mid-2000s that they began to be discussed under the concept of computational journalism (**Thurman**, 2019a). Journalists and specialists in information technology –computer science, statistics, and engineering– began to collaborate with each other by applying computational methods to the search for news stories, and also to the filtering, composition, presentation, and distribution of news. At the same time, the term entered the academic lexicon and academics began to debate the use of computer-based data exploration and interpretation in fulfilling the journalistic role of watchdog. Over time, as journalistic methods diversified, their optimistic view of its potential became more realistic (**Thurman**, 2019b).

When news is produced automatically by computers instead of editors, it is known as automated journalism, a sub-genre of computational journalism (**Thurman**, 2019b), which is also referred to as algorithmic journalism or robot journalism by some authors (**Anderson**, 2013; **Coddington**, 2015; **Graefe**, 2016). In this case, artificial intelligence is used to interpret, organise and present news in a readable format, thanks to algorithms that process huge amounts of data, use pre-programmed news structures, select key aspects and insert names, statistics or images. (**Anderson**, 2013; **Carlson**, 2015; **Graefe**, 2016). In other words, data is converted into narrated news content practically autonomously (**Montal**; **Reich**, 2017) with either minimal or no human intervention (**Carlson**, 2015) through automatic language generation techniques. Thus, what started as a small-scale experiment has become a global phenomenon (**Dörr**, 2016) not only in the production of news stories about sports or economics, for example, but also to complement the work of journalists by suggesting news topics based on trends detected in large databases, or by offering news stories adapted to the needs of different audiences (**Lewis**; **Sanders**; **Carmody**, 2019).

The process described above has given rise to discrepancies around definitions and also to overlapping concepts such as computational journalism, algorithmic journalism or data journalism (**Wright**; **Doyle**, 2019), terms which are often used interchangeably and are difficult to categorise (**Royal**; **Blasingame**, 2016). **Coddington** (2015) contributes to the clarification of this debate and, in summary, highlights three different categories:

- The first, CAR, had its roots in the social sciences and was focused on investigative journalism and public affairs.
- The second, data journalism, is based on the analysis and presentation of data, and is distinguished both by its transparency and interdisciplinarity, and also, like CAR –as **Diakopoulos** (2011) emphasises– by the use of methods borrowed from the social sciences.
- The third category, computational journalism, goes further still, and consists of the application of automation to the news, to which **Graefe** (2016) and **Thurman** (2019b) add that it uses artificial intelligence for the search, selection, verification, composition, presentation and distribution of news stories.

Most of the activities which are integrated into the application of artificial intelligence to journalism have been the subject of review articles that either do not indicate the number of texts analysed or, when they do, are of a number

not usually exceeding a hundred. This is the case with data journalism (**Ausserhofer** *et al.*, 2017; **Zhou**; **Liao**, 2020), the relationship between journalism and technology (**Lewis**; **Westlund**, 2016), computational journalism (**Vállez**; **Codina**, 2018; **Thurman**, 2019b) and automated journalism (**Ali**; **Hassoun**, 2019; **Xu**; **Lan**, 2020). Other works, however, whose remit is not review, provide brief but interesting insights into the literature on automation in journalism (**Túñez-López**; **Toural-Bran**; **Valdiviezo-Abad**, 2019) or touch on other partially related subjects (**Mutsvairo**; **Borges-Rey**; **Bebawi**, 2020; **Meza**, 2016; **Karaboga**; **Karaboga**; **Sehitoglu**, 2020).

2. Aims and research questions

The aim of this paper is threefold. On one hand, to draw a comprehensive map of the scientific production related to the different practices which result from the application of artificial intelligence to journalism; and on the other, to carry out an analysis of this production. The research questions are as follows:

Q1. How has the scientific production on artificial intelligence and journalism developed quantitatively?

Q2. Who are the most prolific authors?

Q3. Which countries produce the most work?

Q4. Is there collaboration between researchers from different countries?

Q5. How have the thematic areas of the field developed and in which direction do they seem to be evolving?

Q6. What research methods and techniques are used?

The answers to these questions provide the basis for the third objective, which is to provide information on under-explored or unexplored areas of study on which other researchers can focus future work.

Q7. What research gaps are identified?

3. Methodology

An analysis of the scientific production on journalism and artificial intelligence was carried out by means of a systematic review, which was not limited to providing a general and quantitative overview of the topic, but which also

> "aimed to gather, evaluate and synthesise the results of primary studies from an integrative perspective" (**Ali**; **Hassoun**, 2019).

Furthermore, in this case, the four criteria or phases that the SALSA (search, appraisal, synthesis and analysis) framework considers necessary for this type of review were applied (**Codina**, 2017), these being: the search for scientific production, the evaluation of results and selection of publications in accordance with certain requirements, the synthesis of results and, finally, the analysis of these results.

Search protocol

Review articles use academic databases such as the *Web of Science* (*WoS*) or *Scopus*, but also *Google Scholar* (among others, **Pearce** *et al.*, 2019; **Thurman**, 2019b) to search for scientific output. The latter has been chosen here because **Martín-Martín** *et al.* (2018) found that it is essentially a combination of *WoS* and *Scopus*, with substantial extra coverage that also includes lower impact papers. This provided a very large starting database from which to subsequently make the necessary eliminations, as suggested by **Prins** *et al.* (2016). **Calvo-Rubio** and **Ufarte-Ruiz** (2021), after studying the use of artificial intelligence in journalism in 209 texts published in the *WoS* and *Scopus* databases, point out precisely the need to expand the sample so as not to exclude equally interesting academic works.

A preliminary search showed that, although there are a considerable number of publications in various languages –mainly Chinese, German and Spanish–, the most widespread of these is English, with many of the others also existing in English versions; so therefore the definitive search for texts was carried out in this language. After several trial runs to make sure the search was as broad as possible, and given the variety of terms related to the application of artificial intelligence to journalism, the following combination of keywords was chosen:

> "robot journalism" OR "computational journalism" OR "automated journalism" OR ("artificial intelligence" AND "journalism") OR ("artificial intelligence" AND "media")

The search was carried out without the application of a time frame and included titles, summaries, keywords, and also full texts, with the anticipation of subsequently discarding the works which were not directly related to the object of study. The time frame of the search was done without a specific start date and with an end date of January 2021, when the search was completed.

This exploration was further complemented by another search strategy (**Schäfer**; **Schlichting**, 2014), which was to manually examine review articles found covering partial aspects of the topic, in order to locate papers that might not have been detected by *Google Scholar*.

Text selection

A first search yielded a total of 980 results ordered by relevance. Of these, those that were not directly related to the topic under study were discarded, such as master's or bachelor's degree theses, press texts, and articles in non-academic journals or blogs. Texts which were considered included articles in indexed journals, book chapters, papers presented at conferences, reports from universities and research centres, and doctoral theses. A total of 351 texts had been collected by January 2021, to which were added 7 from review articles not detected by *Google Scholar*, leading to a final corpus of 358 texts.

Information extraction

After successive modifications, it was established that the following information would be extracted manually after a thorough reading of each text:

- Date of publication.
- Author(s).
- Provenance: country and continent of the university to which each author is affiliated
- Comparative study between countries.
- Type of publication: article, book chapter, conference paper, report, or others.
- Area(s) of application: news detection, automated news verification, robotic writing, data journalism, ethical issues, impact of AI on the textual part of the news, incorporation of AI in journalism education, and others.
- Methodology: quantitative, qualitative, or mixed.
- Research technique: qualitative content analysis, discourse analysis, case study, interviews, focus group, other qualitative, quantitative content analysis, survey, other quantitative, or mixed.
- Assessment of the impact of AI on journalism: positive, negative, or neutral.
- Area of impact: change of model, deontological, economic, teaching, formats, labour, textual, or thematic.
- URL of the text.
- Comments: gist, or most notable facets of the text.

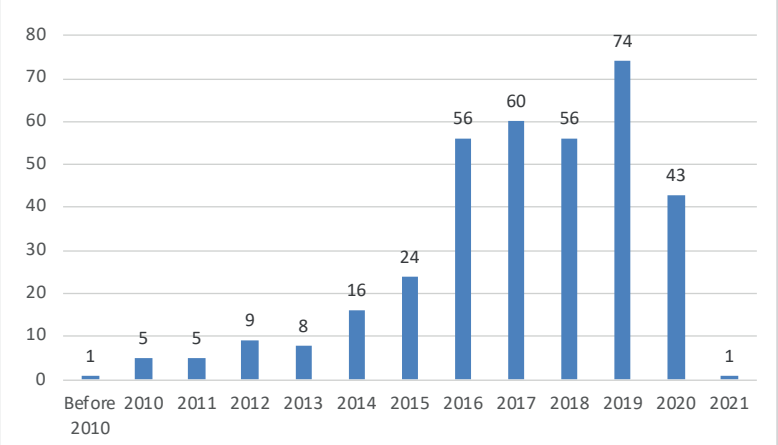
4. Synthesis and interpretation of results

4.1. Temporal and spatial distribution of academic output

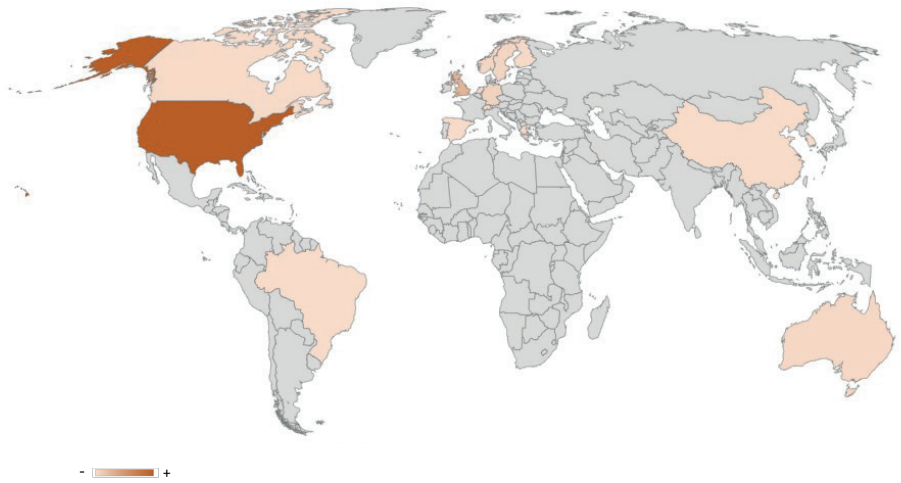
The relationship between artificial intelligence and journalism has been studied with increasing interest over the last decade, especially during the latter half. The results of this study show a notable increase in the academic work produced between the years of 2015 and 2020: of the 358 texts finally reviewed in this study, 87.43% fall within this timeframe. As illustrated in Graph 1, the years 2016 and 2019 constitute two significant turning points. The progression, however, stalls in 2020. The decline over this year will need to be studied in the future to discover whether it really represents a changing tendency or whether, on the contrary, it is just an anomalous result that can be explained by cyclical or exogenous factors, such as the Covid-19 pandemic, for example.

If the origin of the production of academic work is analysed, very significant contrasts can be observed. The current research links each study with the author's university, and with the country where the university is located. Although other procedures could have been used –for example, the countries in which the texts were published–, this approach was chosen because it provides a more revealing map of the research process itself, regardless of circumstances relating to the dissemination or diffusion of these works. As can be seen in Graph 2, the United States and the United Kingdom account for most of the sample analysed, to the extent that these two countries between them have published as many articles (170) as the combined results of all the other countries which have published at least 10 related articles (166 texts). Furthermore, the United States is the clear leader, with three times as many articles published as its nearest rival, the United Kingdom –which in turn is followed at an equally notable distance by a group of Central European countries (Germany, the Netherlands, Switzerland) and Nordic countries (Sweden, Norway, and Finland).

When the countries of the authors' universities are counted, Europe (156 articles) clearly outstrips the Americas (126). What is also revealing here is the appearance of some major contrasts: only three articles from African universities were found, while the output from North Ame-



Graph 1. Time evolution of articles on AI and journalism



Country	N. publications
United States	128
United Kingdom	42
Germany	18
Sweden	16
Norway	15
Australia	14
Netherlands	13
Switzerland	12
Canada	12
Finland	12
Greece	11
Brazil	11
Korea	11
Spain	11
China	10

Graph 2. Distribution of articles by country

rica exceeds one hundred texts (117 in total). South America (9 items) and Oceania (12) are far behind the 38 publications found in Asia. In any case, it should be remembered that this study only considers research published in English, and while English is considered the lingua franca in the field of scientific communication, it is very likely that the results would have been different if papers published in other languages had been examined.

The corpus analysed shows a certain amount of multidisciplinary collaboration. The journalistic point of view needs to be complemented by fields of knowledge as diverse, and often as far apart, as statistics or computer programming. It is possible that in the future, when artificial intelligence applied to journalism has established itself as a distinct discipline –and the gap between journalists and technologists has also narrowed–, true multidisciplinary work will result. So far, task-sharing collaboration has predominated: programmers, for example, have created software tools and journalists have tested or studied their effectiveness (among others, **Schifferes** *et al.*, 2014; **Sim**; **Shin**, 2016; **Tolmie** *et al.*, 2017; **Visvam-Devadoss**; **Thirulokachander**; **Visvam-Devadoss**, 2019). What is clear is that this tenuous collaboration has not yet evolved into any real cooperation between specialists from different countries. The multidisciplinary vision, when it occurs, almost always comes from experts working in the same country. Eighty-six per cent of the texts reviewed are written by authors whose universities are in the same country, and very often these authors belong to the same university. Only 11% of the articles were published by authors doing research in two different countries, while the number of papers involving authors based in three different countries fails to reach 3%.

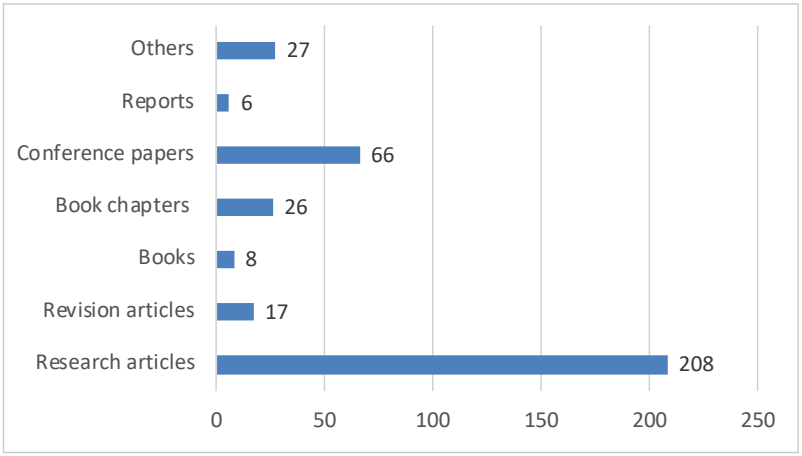
It is not surprising, therefore, that there is a shortage of studies analysing journalistic artificial intelligence projects over multiple countries. Of the 358 studies reviewed, the vast majority (97.2%) refer only to one country. In most cases, the possibility of comparing results from different territories or cultures is not even considered, and a national study is therefore implicitly presented as if it were the only possible option. Only ten articles (2.8% of the sample) present comparative studies involving at least two different countries.

Table 1 lists the names of the authors who appear four or more times throughout the body of texts which were examined for this study. Here again, as per the situation with the nations of origin of the research, there is the double phenomenon of concentration and dispersion: there is a small number of authors with a very high output –the six researchers included in this table account for 15% of the sample analysed– and, following this, a broad spread of authorship, as most signatories have contributed only one paper. This dispersion of authorship could be related to the fact that artificial intelligence is a relatively new discipline –and even newer in its application to journalism. It therefore remains to be seen whether the consolidation of specialised research groups in the field leads to any stability, both quantitative and qualitative, in the production of academic work.

Table 1. Most prolific authors

	First authorship	Second authorship	Third or more
Nicholas Diakopoulos (United States)	13	5	1
Neil Thurman (Germany)	7	1	1
Seth C. Lewis (United States)	7	2	
Ester Appelgren (Sweden)	5		
Eddy Borges-Rey (Qatar)	5		
Meredith Broussard (United States)	4		

4.2. Publication types, research methodologies and techniques



Graph 3. Types of published work

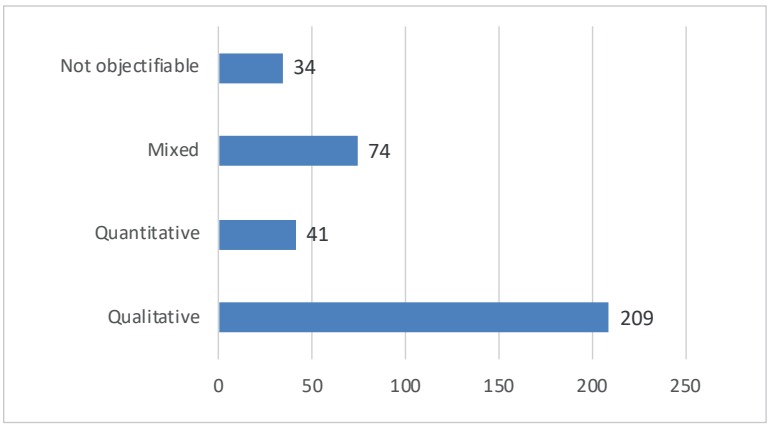
Nearly six out of ten of the works reviewed are articles in academic journals, as shown in Graph 3. However, an analysis of these texts requires some nuances to be introduced and clarified. Both articles and papers –presented at conferences and subsequently published–, which make up 76.50% of this sample, works of an essayistic nature, without either a well-defined methodology or precise research techniques, have a certain presence. This is to some extent to be expected in books or book chapters –and even in some types of reports–, but it is surprising in pure research texts. As previously outlined, this is due to the fact that research in this field has only just begun. As a result, many works reflect generalist, introductory or exploratory approaches. Moving forward towards the current situation, more specific research, with more consistent objectives, methodologies, or developments –and therefore closer to the orthodox research articles usually published in academic journals– has been gaining ground.

Regarding the information contained in Graph 1, it was explained in the previous section that the production of academic work on journalistic artificial intelligence has experienced a notable increase in the last five years, although in 2020 this trend suffered a hiatus. It remains to be seen whether this represents a tendential change or whether it is just the product of the exceptional circumstances of a year marked by the Covid-19 health crisis. It is clear, however, that that this change has particularly affected the publication of articles. The 12 research articles published in 2015 doubled to 24 in 2016; and rose to 32 a year later; while in 2018 and 2019 they rose again –to 35 and 37 articles, respectively. However, in 2020, only 16 articles were published on this subject.

The evolution of other types of publications is very different. Review articles, for example, continued to increase in 2020 (four in 2019, two in 2018, five in 2020). This is perhaps the best indication of the influence exerted by non-academic circumstances: in a social context marked by lockdowns and pandemic prevention measures, ordinary group or team research is very complicated, but the research needed to review the literature already published on this discipline was not affected in the same way. The same is true for chapters taken from books: five books appear in the sample reviewed in 2020, the same number as in 2019 and three more than in 2018.

Graph 4 shows the weight carried by different methodologies through the research into artificial intelligence and journalism. Qualitative approaches are clearly dominant. This prevalence of qualitative research is interesting for several reasons. Firstly, the percentage (close to 60%) that it reaches –much greater than other methodological approaches. Secondly, the very nature of the object of study –close to disciplines such as statistics, database management, machine learning or algorithm development– which would seem to lend itself to a greater prominence of quantitative research. Lastly –and indirectly– the meagre 11.45% which corresponds to works in which quantitative methodology predominates, something which has been notable in recent decades' research into journalism and communication in general. It is also striking that in almost 10% of the publications reviewed it was not even possible to state which methodology had been used. Apart from cases with very specific motivations, this last issue is related to the previously mentioned essay-oriented works, in many of which the methodology used is not mentioned by the authors.

This overview of the methodologies employed can be complemented by a more specific analysis of the specific research techniques used. Table 2 makes the preference for interviews and case studies very clear. Both these techniques are often used in works aiming to investigate to what degree artificial intelligence has been accepted among journalists, who are often reluctant to integrate this kind of innovation into their professional routines. In general, there seems to be a clash between classical journalism –based on humanist conceptions– and the emergence of a new model of journalism dominated by machi-



Graph 4. Predominant methodologies in the reviewed articles

nes, these capable even of writing texts and disseminating personalised news.

Both interviews and case studies have also been used to study the usefulness and popularity of certain tools or applications. In other words, their degree of acceptance or rejection among users. Both techniques are also used in research into the processes of technological innovation in specific media. In these cases, artificial intelligence may be a central element of the research, or only one of many facets studied.

Table 2. Research techniques

Techniques	n	%
Qualitative content analysis	12	3.35
Discourse analysis	3	0.84
Comparative	1	0.28
Case studies	44	12.29
Interviews	49	13.69
Discussion groups	3	0.84
Others (qualitative)	100	27.93
Quantitative content analysis	18	5.03
Surveys (quantitative)	16	4.47
Others (quantitative)	20	5.59
Mixed	53	14.81
Not objectifiable	39	10.89

Table 2 also indicates a minimal, residual use of classic qualitative techniques such as discourse analysis or focus groups. The technique of qualitative content analysis is also in the minority, and given its suitability for studying user opinion, for example, it could theoretically have been a better match for the object of this research. The use of "other" qualitative techniques, as shown in the table above, is found in almost 30% of the research reviewed. This is consistent with pioneering academic work which is still in an early experimental stage of trial and error.

4.3. Application and areas of impact

The scientific literature has considered different facets of artificial intelligence and journalism. Leaving aside the conceptual issues and terminological discrepancies that have already been examined in the first section of this paper, the areas relating to the application of AI in journalism –as reflected in the academic texts compiled for this review article– will be detailed and analysed here. Some areas that are currently in the shadows, but which may soon be fruitfully developed, will also be highlighted in this section.

Table 3 compiles and ranks the main areas identified as priorities by the academic work analysed in this study. As the articles reviewed propose several areas of application, we have chosen to calculate and rank these hierarchically into three groups: principal, secondary and tertiary. According to the body of academic work reviewed, the three principal areas of the application of artificial intelligence in journalism are: data journalism, robotic news writing and news verification. Secondly, and more generally, the impact of new AI technologies on the writing of journalistic texts has also been addressed with some frequency. Similar interest has also been shown in the use of tools that allow information to be extracted and processed –e.g. from social networks– enabling journalists to discover a news event as quickly as possible.

It is worth examining how researchers' interest in each of these areas of application has evolved alongside their development. Some topics, such as the effect of the emergence of artificial intelligence on employment –in terms of job losses and changing models–, have been of largely stable interest. 14 articles were published on these employment issues in 2016, 18 in 2017 and 15 in 2018, indicating no significant change in this area. There is a similar situation regarding the impact of new technology on journalistic writing. However, regarding data journalism, some nuances can already be established: there was a great deal of academic research in this area between 2015 and 2017, but then the growth in the production of work slowed down, and even experienced a considerable decline in 2020. In contrast, robotic writing generated only three academic papers in 2015. In 2018, this number had already risen to 18 articles –a level of focus which was maintained even through 2020.

Evolution in other areas reflects a certain stability. This is the case, for example, with the use of artificial intelligence in verification tasks. Although some isolated articles have been found prior to 2014, it is in 2016 when this issue starts to become addressed regularly. This new interest is influenced not only by aspects related to technological advancement, but also by a different set of political and social circumstances. In 2016, within the framework of the United States presidential campaign, the term "post-truth" gained traction, and was even designated "word of the year" by the *Oxford Dictionary*. Since then, academic attention directed towards this phenomenon has maintained a great regularity. During the period between 2016 and 2020, there are five

Table 3. Areas of use of AI

	Principal	Secondary	Tertiary
Data journalism	111	17	3
Automated news writing	86	12	3
News verification	27	12	2
Textual impact	13	25	1
News detection	25	6	5
AI application on teaching	18	9	4
Ethical problems	13	8	4
Others	65	17	2

academic articles every year (except in 2019, with one text less) whose central interest is the application of artificial intelligence to the verification of news.

> Research on the application of artificial intelligence to journalism has been considerable over the last decade, particularly between the years of 2015 and 2020

The ethical issues arising from the arrival of artificial intelligence in the media have not yet attracted much interest. The peak of academic work related to this issue was reached in 2017 –ten articles–, but the potential for development in this area is still enormous. In the field of writing too, apart from the striking amount of existing work on robotic writing, there is still plenty of research to be done. Not only into new textual procedures –including the models of news personalisation which are now in full effect–, but also into the changes that traditional journalistic writing will have to undergo in order to adapt to this radically different context.

Furthermore, it may be convenient to study the impact of artificial intelligence on journalism from an even broader perspective: in the coming years it will continue to be necessary to describe and analyze the classic tasks that journalists encounter (information searches, data verification, hierarchical organization of content, or news writing), but now relating them to new tasks that affect the growing interaction between man and machine. In this sense, it is clear that it will be essential to reformulate the professional routines of journalists in aspects as fundamental as the detection of relevant facts, the management of large databases, the journalistic use of social media or the verification of content. As a result, academic studies must also investigate the new role of journalists in a newsroom that is no longer disconnected from the influence of AI.

The analysis of works that link the professional and academic realms allows us to predict an intensification in the amount of research which reviews the training needs of journalists –both current and future. The studies published so far argue that it is essential and urgent to adapt university curricula to such new realities as the effects and use of artificial intelligence in journalism. Furthermore, it is important that the university teaching staff themselves undergo renewal and updating of their knowledge, otherwise it will be impossible for them to inculcate in their students the competences and skills that the media today are increasingly demanding. In the same vein, it is very interesting to see how journalists who left university years ago –many of whom are now in senior positions– are adapting to working environments in which the weight of technology is infinitely greater than it was at the beginning of this century. This process of adaptation, with the resulting tension between journalists and technologists, has already been experienced during the processes of digitalisation of newsrooms and the emergence of social networks. It is therefore foreseeable that something similar will now occur with the rise of artificial intelligence.

A positive view of the effects of the advent of artificial intelligence in the media has predominated in the research papers reviewed to date. In fact, almost no articles exist which voice a negative view towards these new technologies. In academic texts a neutral attitude is now more usually found, this being the case in almost 60% of the works analysed. It is interesting to observe how this neutral attitude has evolved over the last decade: in the early years of the previous decade, a colder and more distanced outlook was quite apparent: in 2012, for example, 66.6% of the texts recorded were neutral. Three years later, in 2015, this neutral view had faded considerably, appearing in only 50% of the research. Conversely, the number of articles iterating a more positive and less critical view of the effects of artificial intelligence on journalism had increased. However, in the last few years studied, this positive view has weakened and there has been a return to neutrality, with rates of around 60%, and even –as recorded in 2018– close to 65%.

5. Discussion and conclusions

Research on the application of artificial intelligence to journalism is still in its infancy, which is logical when considering that the emergence of the use of algorithms and artificial intelligence in the profession is also relatively new. However, the interest in this phenomenon, which has already changed many journalistic practices and areas, has resulted in some interesting academic work which seems to be coalescing into different, and better defined, areas of study.

This review paper charts an overall map of the scientific work published on this subject, and describes and analyses how it has developed: its evolution over time, which countries have published the most, which areas have received the most attention, which authors are the most prolific, what assessment is made of the phenomenon, what type of publication is the most plentiful, and which methodologies and research techniques are the most used. Added to this, some gaps in the scientific literature published to date are pointed out, which may serve to focus future research.

In contrast to other review articles, an attempt was made here to gather as broad a body of work as possible and to include all the areas that have been the subject of study. It should be remembered that the published articles do not exceed one hundred reviewed texts. Some authors do not specify the number of texts reviewed (such as **Ausserhofer** *et al.*, 2017; **Zhou**; **Liao**, 2020; **Lewis**; **Westlund**, 2016; **Vállez**; **Codina**, 2018; **Thurman**, 2019b; **Ali**; **Hassoun**, 2019; **Xu**; **Lan**, 2020), others do not have a review as their main objective (**Túñez-López**; **Toural-Bran**; **Valdiviezo-Abad**, 2019) and others focus only on partial aspects (**Mutsvairo**; **Borges-Rey**; **Bebawi**, 2020; **Meza**, 2016; **Karaboga**;

> The United States clearly leads the ranking of countries with the most publications, and triples the production of the United Kingdom

Karaboga; **Sehitoglu**, 2020).

> Data journalism, robotic news writing and news verification have been the subject of the most academic studies

After analysing the 358 texts collected for this review, it is clear that academic interest in the application of artificial intelligence to journalism has been considerable over the last decade, particularly between the years of 2015 and 2020, a period in which 87.43% of the articles reviewed here are concentrated. Regarding their geographical origin, there are revealing contrasts: The United States (with 128) clearly leads the ranking of countries with the most publications and triples the production of the United Kingdom, which is in second place, with 42. The research by **Calvo-Rubio** and **Ufarte-Ruiz** (2021) cited above reaches similar conclusions. Following the United States and the United Kingdom, at a considerable distance, are a group of Central European countries (Germany, the Netherlands and Switzerland) and Nordic countries (Sweden, Norway, and Finland) with 43 publications between them. At this point, it should be remembered that the selected corpus only includes articles published in English, as this is considered the international language for scientific research. Although we cannot know whether the results would have been the same if this criterion had not been applied to the selection of the sample, it is also true that many texts written in other languages also have an English version. It is becoming increasingly common for academic journals to include an English version, in addition to the text in its original language.

The limited collaboration between researchers from different countries –86% of the texts reviewed are co-written by authors from universities located in the same country– may explain the few comparative studies of how artificial intelligence is applied to journalism in different places –only 2.8% of the sample–. There is also a very limited amount of research by specialists from different scientific disciplines, which is particularly surprising in the analysis of a phenomenon that seems to require constant cooperation and dialogue between journalists and professionals from other fields, as is pointed out by some authors (**Thurman**, 2019a).

As might be expected in a new subject area, there is considerable methodological variety and vagueness, although most of the articles (58%) use a qualitative method and rely mainly on focus groups and interviews as research techniques. The new challenges and areas of study that emerge in the future, however, will also require new research methods. Neither are there many well-established research groups: only six of the authors of the 358 texts reviewed in the current study have published four or more articles on this topic.

Interest in the impact of intelligent machines on journalism crosses into different areas, most of which are crucial for the journalistic models of media companies. Of these, data journalism (111), robotic news writing (86) and news verification (27) have been the subject of the most academic studies. There are, however, many others that have not yet received the same attention, such as the necessary rethinking of the journalist´s role, the possibility of personalising content for specific audiences, or the training needs in both the academic and private sectors, but interest in these areas will surely increase soon. As the application of artificial intelligence to different journalistic practices spreads and develops, the interaction between journalists and machines will increase, and thus it will be necessary to reformulate the organisation of work in newsrooms. Academics will also have to investigate, thus, these new roles in an activity marked by the determining influence of AI.

Everything seems to indicate that these changes are the beginning of what may in the next few years become one of the most radical transformations that journalism has ever experienced, as it will affect business models and the very ways in which texts are disseminated, as well as impacting on professional roles and practices. After the digitisation of newsrooms and publishing media, the emergence and integration of social networks in the production and distribution of journalistic texts, and the problems of disinformation and fake news, it is foreseeable that the next phase of technological revolution will be brought about by the emergence of artificial intelligence.

Most of the texts reviewed for this article do not offer a negative view of the application of artificial intelligence to journalism. After a period (2015-2016) in which a very favourable and uncritical attitude was shown, the most recently published texts (2017-2019) return to the neutral tone which was common in the earlier years of the last decade (2011-2012). Some of them even provide new critical approaches in some areas, which may represent a certain level of maturity and consolidation in this discipline.

As this new phenomenon becomes established in the journalistic profession and gains a foothold in academic research, universities face a major challenge in adapting their teaching to the radical changes it brings about. The publication of academic studies could help in this respect: for example, it would be interesting to study whether the countries with the highest academic output on this topic are the same as those whose media are making the most use of artificial intelligence, or whose universities have incorporated it into their curricula.

> The journalist's role, the possibility of personalising content for specific audiences, or the training needs in both the academic and private sectors are aspects that need to be more studied

6. References

Aghion, Philippe; **Jones, Benjamin F.**; **Jones, Charles I.** (2019). "Artificial intelligence and economic growth". In:

Agrawal, Ajay; Gans, Joshua; Goldfarb, Avi. *The economics of artificial intelligence*. Chicago: University of Chicago Press, pp. 237-290.
https://press.uchicago.edu/ucp/books/book/chicago/E/bo35780726.html

Ali, Waleed; **Hassoun, Mohamed** (2019). "Artificial intelligence and automated journalism: Contemporary challenges and new opportunities". *International journal of media, journalism and mass communications*, v. 5, n. 1, pp. 40-49.
https://doi.org/10.20431/2454-9479.0501004

Anderson, Christopher W. (2013). "Towards a sociology of computational and algorithmic journalism". *New media and society*, v. 15, n. 7, pp. 1005-1021.
https://doi.org/10.1177/1461444812465137

Ausserhofer, Julian; **Gutounig, Robert**; **Oppermann, Michael**; **Matiasek, Sarah**; **Goldgruber, Eva** (2017). "The datafication of data journalism scholarship: focal points, methods, and research propositions for the investigation of data-intensive newswork". *Journalism*, v. 21, n. 7, pp. 950-973.
https://doi.org/10.1177/1464884917700667

Bounegru, Liliana (2012). "Data journalism and computer-assisted reporting". In: Gray, Jonathan; Chambers, Lucy; Bounegru, Liliana. *The data journalism handbook: How journalists can use data to improve the news*. Sebastopol, CA: O'Reilly, pp. 21-22. ISBN: 978 1 449 33006

Calvo-Rubio, Luis-Mauricio; **Ufarte-Ruiz, María-José** (2021). "Artificial intelligence and journalism: Systematic review of scientific production in Web of Science and Scopus (2008-2019)". *Communication & society*, v. 34, n. 2, pp. 159-176.
https://doi.org/10.15581/003.34.2.159-176

Carlson, Matt (2015). "The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority". *Digital journalism*, v. 3, n. 3, pp. 416-431.
https://doi.org/10.1080/21670811.2014.976412

Coddington, Mark (2015). "Clarifying journalism's quantitative turn". *Digital journalism*, v. 3, n. 3, pp. 331-348.
https://doi.org/10.1080/21670811.2014.976400

Codina, Lluís (2017). *Revisiones sistematizadas y cómo llevarlas a cabo con garantías: systematic reviews y Salsa framework*.
https://www.lluiscodina.com/revision-sistematica-salsa-framework

Cox, Melisma (2000). ''The development of computer-assisted reporting". In: *Newspaper Division, Association for Education in Journalism and Mass Communication*, pp. 1-22.
https://cutt.ly/YkMPYNx

De-Lima-Santos, Mathias-Felipe; **Schapals, Aljosha-Karim**; **Bruns, Axel** (2020). "Out-of-the-box versus in-house tools: how are they affecting data journalism in Australia?". *Media international Australia*, pp. 1-15.
https://doi.org/10.1177/1329878X20961569

Diakopoulos, Nicholas (2011). *A functional roadmap for innovation in computational journalism*.
https://cutt.ly/LkMd22h

Dörr, Konstantin-Nicholas (2016). "Mapping the field of algorithmic journalism". *Digital journalism*, v. 4, n. 6, pp. 700-722.
https://doi.org/10.1080/21670811.2015.1096748

Graefe, Andreas (2016). "Guide to automated journalism". *Columbia journalism review*, 7 January.
https://cutt.ly/SkMufyO

Henn, Steve (2012). "The night a computer predicted the next president". *NPR*, 31 October.
https://cutt.ly/BkMQU1N

Karaboğa, Tuğba; **Karaboğa, Hasan-Aykut**; **Şehitoğlu, Yasin** (2020). "The rise of big data in communication sciences: a bibliometric mapping of the literature". *Connectist: Istanbul University journal of communication sciences*, v. 58, pp. 169-199.
https://doi.org/10.26650/CONNECTIST2020-0083

Léchenet, Alexandre (2014). *Global database investigations: The role of the computer-assisted reporter*. Reuters Institute fellowship paper.
https://cutt.ly/FkNG99C

Lewis, Seth C.; **Sanders, Amy-Kristin**; **Carmody, Casey** (2019). "Libel by algorithm? Automated journalism and the threat of legal liability". *Journalism & mass communication quarterly*, v. 96, n. 1, pp. 60-81.
https://doi.org/10.1177/1077699018755983

Lewis, Seth C.; **Westlund, Oscar** (2016). "Mapping the human-machine divide in journalism". In: Witschge, Tamara; Anderson, C. W.; Domingo, David; Hermida, Alfred. *The SAGE handbook of digital journalism*. London: Sage, pp. 341-353. ISBN: 978 1 4739 0653 2

Martín-Martín, Alberto; **Orduña-Malea, Enrique**; **Thelwall, Mike**; **Delgado-López-Cózar, Emilio** (2018). "Google Scholar, Web of Science, and Scopus: a systematic comparison of citations in 252 subject categories". *Journal of informetrics*, v. 12, n. 4, pp. 1160-1177.
https://doi.org/10.1016/J.JOI.2018.09.002

McGregor, Susan (2013). "CAR hits the mainstream". *Columbia journalism review*, 18 March.
http://www.cjr.org/data_points/computer_assisted_reporting.php

Meza, Radu (2016). "Computational thinking and journalism education". In: Singla, Carles; Da-Rocha, Irene; Ramon, Xavier. *Shaping the future of news media. The international conference on integrated journalism education, research and innovation proceedings.* Barcelona: Integrated Journalism in Europe, pp. 21-22. ISBN: 978 84 606 9570 7
https://cutt.ly/ckMPLlw

Montal, Tal; **Reich, Zvi** (2017). "I, robot. You, journalist. Who is the author?". *Digital journalism*, v. 5, n. 7, pp. 829-849.
https://doi.org/10.1080/21670811.2016.1209083

Mutsvairo, Bruce; **Borges-Rey, Eddy**; **Bebawi, Saba** (2020). *Data journalism in the Global South*. Cham: Springer. ISBN: 978 1 4739 0653 2

Pearce, Warren; **Niederer, Sabina**; **Özkula**; **Suay-Melisa**; **Sánchez-Querubín, Natalia** (2019). "The social media life of climate change: Platforms, publics, and future imaginarie". *WIREs clim change*, v. 10, p. 2, e569.
https://doi.org/10.1002/wcc.569

Prins, Ad A. M.; **Costas, Rodrigo**; **Van-Leeuwen, Thed N.**; **Wouters, Paul F.** (2016). "Using Google Scholar in research evaluation of humanities and social science programs: A comparison with Web of Science data". *Research evaluation*, v. 25, n. 3, pp. 264-270.
https://doi.org/10.1093/reseval/rvv049

Royal, Cindy; **Blasingame, Dale** (2015). "Data journalism: An explication". *ISOJ*, v. 6, n. 1.
https://isojjournal.wordpress.com/2015/04/15/data-journalism-an-explication

Schäfer, Mike S.; **Schlichting, Inga** (2014). "Media representations of climate change: A meta-analysis of the research field". *Environmental communication*, v. 8, n. 2, pp. 142-160.
https://doi.org/10.1080/17524032.2014.914050

Schifferes, Steve; **Newman, Nic**; **Thurman, Neil**; **Corney, David**; **Göker, Ayse**; **Martin, Carlos** (2014). "Identifying and verifying news through social media". *Digital journalism*, v. 2, n. 3, pp. 406-418.
https://doi.org/10.1080/21670811.2014.892747

Sim, Da-Hun; **Shin, Seung-Jung** (2016). "Implementation of algorithm to write articles by stock robot". *International journal of advanced smart convergence*, v. 5, n. 4, pp. 40-47.
https://doi.org/10.7236/IJASC.2016.5.4.40

Thurman, Neil (2019a). "Personalization of news". In: Vos, Tim; Hanusch, Folker. *The international encyclopedia of journalism studies*. Hoboken, NJ: John Wiley & Sons, pp. 1-6.
https://doi.org/10.1002/9781118841570.iejs0052

Thurman, Neil (2019b). "Computational journalism". In: Wahl-Jorgensen, Karin; Hanitzsch, Thomas. *The handbook of journalism studies*, New York: Routledge, pp. 180-195.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3265967

Tolmie, Peter; **Procter, Rob**; **Randall, David-William**; **Rouncefield, Mark**; **Burger, Christian**; **Wong-Sak-Hoi, Geraldine**; **Zubiaga, Arkaitz**; **Liakata, Maria** (2017). "Supporting the use of user generated content in journalistic practice". In: *CHI '17: Proceedings of the 2017 CHI conference on human factors in computing systems*, pp. 3632-3644.
https://doi.org/10.1145/3025453.3025892

Túñez-López, José-Miguel; **Toural-Bran, Carlos**; **Valdiviezo-Abad, Cesibel** (2019). "Automation, bots and algorithms in newsmaking. Impact and quality of artificial journalism". *Revista latina de comunicación social*, v. 74, pp. 1411-1433.
https://doi.org/10.4185/RLCS-2019-1391en

Vállez, Mari; **Codina, Lluís** (2018). "Periodismo computacional: evolución, casos y herramientas". *El profesional de la información*, v. 27, n. 4, pp. 759-768.
https://doi.org/10.3145/epi.2018.jul.05

Visvam-Devadoss, Ambeth-Kumar; **Thirulokachander, Vijay-Rajasekar**; **Visvam Devadoss, Ashok-Kumar** (2019). "Efficient daily news platform generation using natural language processing". *International journal of information technology*, v. 11, pp. 295-311.
https://doi.org/10.1007/s41870-018-0239-4

Weber, Wibke; **Engebretsen, Martin**; **Kennedy, Helen** (2018). "Data stories. Rethinking journalistic storytelling in the context of data journalism". *Studies in communication sciences*, v. 18, n. 1, pp. 191-206.
https://doi.org/10.24434/j.scoms.2018.01.013

Wright, Scott; **Doyle, Kim** (2019). "The evolution of data journalism: A case study of Australia". *Journalism studies*, v. 20, n. 13, pp. 1811-1827.
https://doi.org/10.1080/1461670X.2018.1539343

Xu, Zhichao; **Lan, Xintong** (2020). "A scientometric review of automated journalism: Analysis and visualization". *Journal of physics: Conference series*, v. 1684, 012127.
https://doi.org/10.1088/1742-6596/1684/1/012127

Zhou, Yujin; **Liao, Han-Teng** (2020). "A bibliometric analysis of communication research on artificial intelligence and big data". In: *6th International conference on humanities and social science research (Ichssr 2020)*, pp. 456-459.
https://doi.org/10.2991/assehr.k.200428.097