
Abstract
Research about the use of Artificial Intelligence applied to journalism has increased over the years. The studies conducted in this field between January 2008 and December 2019 were analysed to understand the contexts in which they have been developed and the challenges detected. The method used consisted of a systematic review of the scientific literature (SLR) of 209 scientific documents published in the Web of Science and Scopus databases. The validation required the inclusion and exclusion criteria, database identification, search engines and evaluation and description of results. The findings indicate that the largest number of publications related to this topic are concentrated in the United States and that the rise of scientific production on Artificial Intelligence in journalism takes place in 2015, when the remarkable growth of these publications begins, until reaching 61 in 2019. It is concluded that research is mainly published in scientific journals, which include works that handle a broad variety of topics, such as information production, data journalism, big data, application in social networks or information checking. In relation to authorship, the trend is the presence of a single signer.

Keywords
Artificial Intelligence, systematized bibliographic review, scientific production, Web of Science, Scopus, journalism.

1. Introduction
The scientific interest of the impact of Artificial Intelligence (AI onwards) in journalism has increased over the years (van der Kaa & Krahmer, 2014). From the first studies about the subject (Kim et al., 2007; Macau, 2004; Matsumoto et al., 2007) and up until the latest contributions (Calvo Rubio & Ufarte, 2020; Dörr, 2016; Fernández-Torres, Gutiérrez-Fernández & Palomo-Zurdo, 2019; Hansen et al., 2017; Oppenheimer, 2018; Salazar, 2018, and Ufarte, Túñez & Vaz, 2019, among others), it was demonstrated that the media industry is facing the challenge of contents robotisation (Murcia & Ufarte, 2019; Sandoval et al., 2019; Usher, 2017; Weeks, 2014), referring to the algorithmic process that turns data into narrative and informative texts with little or null human intervention, apart from the initial programming (Barrat, 2013; Bunz, 2010; Harcup, 2014).

Media have not hesitated in using this process to mechanise search and classification functions (Lemelshtrich, 2018; Lindén, 2017). The first experiences produced in the United
States back in 2014, by *Los Angeles Times* (Flores-Vivar, 2018) and *Associated Press* (Brandon, 2014; Lichterman, 2017). Soon, media from Brazil (Monnerat, 2018) followed, as well as media from China (Martín, 2017), Japan (Jung et al., 2017), United Kingdom (Gani & Haddou, 2014), Finland (Melin et al., 2018), Germany (Horky & Pelka, 2017), France (Sánchez & Sánchez, 2017), Sweden (Stern, 2017) and Spain (Southern, 2017), as well as communication agencies from the Netherlands, Austria, Denmark, Portugal, Norway, and Sweden (Fanta, 2017). Túñez, Toural and Cacheiro (2018) counted a total of 16 journalistic media, 13 news agencies and 21 companies that used the automation of journalistic texts.

The arguments for introducing robots on editorial offices have been many: greater accuracy (Silverman, 2013), increase of production (Papadimitriou, 2016), objectivity (Graefe, 2016), the ability to add web content (Mittal & Kumaraguru, 2014; Starbird et al., 2010), customising information (Hwang, Pearce & Nánis, 2012; Newman et al., 2010; Slater & Rouner, 2002; Young & Hermida, 2015), detecting events of journalistic interest for later diffusion (Steiner, 2014) and fighting against disinformation (Flew, 2012), especially in news and journalistic briefs (Ufarte & Manfredi, 2019), but also in Tweets and graphics (Rojas Torrijos & Toural, 2019). Likewise, they are used to identify the management of profiles on social networks (Chu et al., 2010; Dickerson, Kagan & Subrahmanian, 2014; Ferrara et al., 2016; Tavares & Faisal, 2013) and adapt to the preferences of users (Keeney, 2015).

But the use of algorithms has generated opposite reactions. Strömbäck (2005) considers that it entails a rupture with the idea of what journalism means, not only because these programmes cannot formulate questions and issue opinions, but because they may be inadequate to play a *watchdog* role in a disperse environment. On the other hand, Bostrom (2014); Bravo Orellana, Santana Ormeño and Rodón (2014); Brynjolfsson and McAfee (2014); Cid (2017), Valdiviezo and Bonini (2019) and Villoro (2015) say that AI can cause an unemployment crisis in the journalistic sector, because the algorithm simulates the natural language through a software that creates computerised texts with the same characteristic of those written by a human (Lokot & Diakopoulos, 2016). However, authors like Acemoglu and Restrepo (2019); Cerezo (2018), Cervera (2017); Cosoy (2017), Renó (2018), Salaverría (2016) and Túñez and Toural (2018) advocate there is not a real danger of extinction for the profession, but instead a process of changes and adjustments, because this technology is a driving force of growth for organisations (David, 2015; Mark, 2019), therefore, adaption is a priority that demands knowing its features and singularities (Powers, 2012).

The increase of research demonstrates that the application of AI to journalism is a growing field of interest characterised by transdisciplinarity. Nevertheless, this trend is not always transferred with the same intensity to all areas of knowledge. This happens, with the meta–research about automation in journalism, understood as the research about the works performed and published in this field of study (Barranquero & Ángel, 2015; Roca & Pueyo, 2012; Rodríguez Gómez, Goyanes & Rosique, 2018). This discipline made its first steps at the beginning of the century (Jones, 2007; Moragas, 2005), and in recent years it had regained protagonism (Baladrón, Correyero & Manchado, 2014; Caffarel, 2018) thanks to the studies of Caffarel, Ortega Mohedano and Gaitán (2017), Delgado López Cózar and Martín Martín (2015), Fernández Quijada and Masip (2013) and Navarro Beltrá and Martín Llaguno (2013), among others, who analyse the scientific production in communication. In this same line, other specific studies in journalism (Martínez Nicolás, Saperas & Carrasco Campos, 2017), the mobile journalism (López et al., 2019), fake news (Blanco Alfonso, García Galera & Tejedor, 2019; Valero & Oliveira, 2018), transmedia narrative (Vicente Torrico, 2017), big data (Gindin, Busso & 2018), digital competencies in journalism (Marta–Lazo, Rodríguez & Peñaña, 2020) or the role of women in communication research (Martín Algarra, Serrano Puche & Rebollero, 2018) are noteworthy.

On the contrary, there are no studies that analyse publications about the use of AI in the journalistic sector. In the national scope, only the study of Túñez, Toural and Valdiviezo (2018)
offers an analysis, but without detailing descriptive nor content data. Therefore, it is necessary to review the specialised literature, in order to know under what terms the academic discussion is developed. The purpose is making a static picture that consolidates the progresses registered in said field (Guardiola, 1991) and that serves to orientate future researches (Guirao, 2015). The aim is providing knowledge and understanding the complexity of this technology (Eudes, 2014; Newman, 2018).

Therefore, this research is timely and appropriate, not only due to its theoretical, bibliometric and exploratory value, but because it fills a gap that persists in this field of study when answering the following research questions:

Q1. When the researches focused on AI in journalism have hatched?
Q2. What are the leading countries?
Q3. What is the authors’ affiliation?
Q4. What sort of publications concentrate more works about this subject?
Q5. What is the authorship of the researches about robot journalism?
Q6. What is the approach used in the researches?
Q7. What subjects are the analysed and studied the most?

2. Methodology

2.1. Objectives and method

The objective of this study is to analyse the recent works (2008–2019) about the application of AI in journalism. A chronological research is proposed to demonstrate the validity of this field of study; a territorial exploration, to discover the main countries of publication; an institutional review to confirm the most prolific universities in this area; a bibliographic analysis, to know what publications shelter the scientific production; an analysis of authors to understand the creative synergies and, lastly, a content analysis, to identify what are the themes treated the most to establish the main research lines.

The methodological design used lies in the protocol-based systematic scientific literature review (SLR) proposed by Kitchenham (2004). This technique is part of a secondary research. Its basics are the respect for transparency and systematisation (Tranfield, Denyer & Smart, 2003) and facilitates the identification, contrast, assessment and interpretation of relevant researches (Codina, 2017). Furthermore, it allows to identify, evaluate and interpret available data in a determined field of research and within a specific period of time (Ramírez Montoya & García Peñalva, 2018).

This method describes the intellectual story of a discipline (Pasadeos, Phelps & Bong-Hyun, 1998) and it is very useful in areas under constant evolution (Cué et al., 1996; Guirao, Olmedo & Ferrer, 2008), like the AI in journalism, because it provides a background for managing the increase of publications in a determined area thanks to the use of a standardised and reproducible protocol that guarantees quality, clarity and consistency of the review process (Coughlan, Cronin & Ryan, 2013). The studies oriented to analysing the dominant trends of research are a symptom of maturity for any scientific discipline (Martínez Nicolás, 2009) and have turned into measuring instruments, accepted, and recognised by the international scientific community (Bordons, Fernández Bajón & Gómez Caridad, 2002; Waltman et al., 2012).

2.2. Analysis period and selection of sources

This longitudinal research analyses an 11 years period: from 1 January 2008 until 31 December 2019. We have decided for 2008 as the first year of the period because it was then when the definitive onset of AI in the journalistic sector took place (Salazar, 2018) to satisfy the wide demand of information and to reduce the production cost (Podolny, 2015).
Regarding sources, the databases of Web of Science (WOS) and Scopus were selected. These are the two main scholar international information sources (Maltrás, 2003), and a suitable resource to conduct this sort of studies (Codina, Olmeda Gómez & Perianes, 2013).

2.3. Automated search using descriptors

Once the sources were delimited, a search strategy was designed to gather researches related with the study object. This process started with an exploratory screening, where keywords referring to the use of AI in journalism were used. The search descriptors were the different denominations in English used by researchers to describe this way of writing journalistic texts, since there is no unanimity in the conceptualisation of this phenomenon in the Academy. Specifically, there were eight terms selected: 1) Robot journalism, a name established by Burrell (2016), Carlson (2015), Dawson (2010), Lee and Kim (1998), Levy (2012), Lindén (2017), Matsumoto et al. (2007), Montal and Reich (2016), Oremus (2015) and Thurman, Dörr and Kunert (2017); 2) Algorithm journalism, established by Anderson (2013), Bercovici (2010), Diakopoulos (2019) and van Dalen (2012); 3) Automated journalism, used by Caswell and Dörr (2018), Napoli (2014) and Wölker and Powell (2021); 4) Computational journalism, indicated by Clerwall (2014), Coddington (2015), Cohen, Hamilton and Turner (2011), Gynnild (2014), Karlsen and Stavelin (2014) and Váquez and Codina (2018); 5) Augmented journalism, designated by Ferrer Conill, (2015), Marconi and Siegman (2017) and Pavlik and Bridges (2013); 6) Artificial journalism, a name granted by Túñez, Toural and Valdiviezo (2019); and 7) High tech journalism, indicated by Salaverría (2014). The eight descriptor responds to the combined search of the terms journalism and artificial intelligence.

The results in this first stage of search were 481 registries: 230 in WOS and 251 in Scopus. These publications were reviewed to eliminate duplicated entries and false retrievals, discarding all entries not related to the theme of analysis (Table 1). The printed and digital editions of the different journals were considered as one (Carpenter, 2008) and only those publications of journalistic themes were selected. As final step, all entries that did not contain all the data have been reviewed individually and the works where no documentary support could be found where discarded (4).

Table 1: Descriptors selected for the search in the Web of Science and Scopus.

<table>
<thead>
<tr>
<th>Descriptors</th>
<th>Denomination in Spanish</th>
<th>Denomination in English</th>
<th>Number of publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Periodismo robot</td>
<td>Robot journalism</td>
<td>30</td>
<td>24</td>
</tr>
<tr>
<td>Periodismo algorítmico</td>
<td>Algorithm journalism</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Periodismo automatizado</td>
<td>Automated journalism</td>
<td>34</td>
<td>39</td>
</tr>
<tr>
<td>Periodismo computacional</td>
<td>Computational journalism</td>
<td>98</td>
<td>98</td>
</tr>
<tr>
<td>Periodismo aumentado</td>
<td>Augmented journalism</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Periodismo artificial</td>
<td>Artificial journalism</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Periodismo de alta tecnología</td>
<td>High tech journalism</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Periodismo + Inteligencia Artificial</td>
<td>Journalism and Artificial Intelligence</td>
<td>62</td>
<td>89</td>
</tr>
<tr>
<td>Total in databases</td>
<td></td>
<td>230</td>
<td>251</td>
</tr>
</tbody>
</table>

Source: Own elaboration.

The search was done in July 2019 and has been updated in two occasions, in September 2019 and February 2020.
2.4. Identification of study variables

The analysis protocol is based on authors like López Rabadán and Vicente Mariño (2011), Barranquero and Marín García (2014), Galán et al. (2018) or López Robles et al. (2019).

A coding card was applied to each analysis unit and composed of different variables in compliance with every objective aimed in the study. Firstly, this coding card includes the descriptive publication data, such as the year, the journal’s name, the title of the research, number of authors, affiliation centre and the country of the first signer. Secondly, the theme studied, and the keywords were analysed (Table 2).

For the theme affiliation, the keywords related to the terms used for the article search were not used (computational journalism, automated journalism, robot journalism, journalism, artificial intelligence, algorithms, automation, algorithmic journalism, robots), as well as other generic concepts (innovation, technology, digital journalism, news, journalism innovation, journalism studies, capital, digital media, online journalism, internet, media, newspaper). The remaining concepts in each one of the articles have been associated with the following themes: production, data journalism, big data, theory, social media, fact-checking, technologies, personalisation, audience, credibility, ethics, events detection, online comments, authorship and chatbots.

In addition, the research approach was added to the variables analysed. This variable was included using these values: exploratory, descriptive, correlational, explanatory, and experimental.

For classification into these categories, the definitions provided by Hernández Sampieri, Fernández Collado and Baptista Lucio (2010) have been considered. These authors believe that an exploratory research is the one analysing a scarcely studied problem analysing innovative phenomena. Moreover, there are the descriptive works, through information about concepts or variables, looking for “specifying the properties, characteristics and profiles of people, groups, communities, processes, objects or any other phenomenon” (p. 80). The researches with a correlational approach aim to “understand the relationship of the level of association between two or more concepts, categories or variables in a particular context” (p. 81). Lastly, researchers consider that studies of explanatory reach are more structured and are targeted to knowing “why a phenomenon takes place and under what conditions it manifest, or why two or more variables associate with each other” (p. 84). After a first approach to the articles found in the databases, we decided to extend this classification with the researches using an experimental design to refer to those works aimed to testing the use of tools linked to AI.

Furthermore, the reach of publications has been assessed through data of citations.

Table 2: Study variables.

<table>
<thead>
<tr>
<th>Descriptive data</th>
<th>Publication year</th>
<th>Journal’s name</th>
<th>Research title</th>
<th>Number of authors</th>
<th>Institution of origin (first author)</th>
<th>Country (first author)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thematic field</td>
<td>Theme</td>
<td>Keywords (in English)</td>
<td>Research approach</td>
<td>Reach</td>
<td>Total of citations</td>
<td></td>
</tr>
</tbody>
</table>

Source: Own elaboration.

To guarantee the reliability of coding, the application of the coding card to every publication has been made in parallel by the researchers, previously defining the analysis guidelines (double-check), reviewing the possible contradictions in preceding meetings. This stage was
essential for the theme analysis, of a high complexity due to the great variety of approaches. The results obtained have been analysed with the Excel Microsoft programme, by means of the frequency count and the crossing of variables.

3. Results

3.1. Publication years

209 publications were reviewed, a result of the search of eight terms or set of terms and the elimination of possible coincidences and false positives (Table 1). Specifically, 129 scientific articles have been published (represent 61.7% from the total), 58 conference papers (27.8%), 12 book chapters (5.7%), 3 press articles (1.4%), 2 books (1%), 2 book introductions (1%), 2 reviews (1%) and 1 Workshop report (0.5%).

Regarding the publication date, it is confirmed that the academic literature on this matter has been in crescendo in the past decade, due to a progressive increase in the production of AI-related contents in journalism, reaching a peak in 2019, with 61 publications associated with the selected descriptors (Figure 1).

![Figure 1: Annual production of publications about AI in journalism.](source)

In 2015, a turning point was observed, in such a way that the last five years, hoard 85.2% of the scientific production. In these five years, the research articles have grown from 10 in 2015 until 44 in 2019; being 11 in 2016, 23 in 2017 and 29 in 2018. Therefore, a clear increasing trend is observed. However, this does not happen with the conference papers, showing an erratic evolution: 6 in 2015, 9 in 2016, 8 in 2017, 10 in 2018 and 6 in 2019. In this period all the production is accumulated in book format, like complete volumes, chapters, and introductions.

3.2. Outstanding countries and universities or research centres

The researchers signing the 209 publications analysed come from institutions located in 31 different countries. Among all of them, the presence of centres in the United States stand out, gathering 33% (N=69) of the corpus reviewed. South Korea (N=19) is the second country with the greatest production on this study subject, with 9.1% of the total, followed by Spain (N=15) and Germany (N=15), both with 7.2%. Followed by Great Britain (N=9), Ireland (N=7), Israel (N=6); Canada (N=5); Brazil (N=5), Russia (N=5), Greece (N=5), Slovenia (N=4), India (N=4), Singapore (N=4), Norway (N=4), Australia (N=3), China (N=3), Denmark (N=3), Finland (N=3), Netherlands (N=3), Portugal (N=3), Switzerland (N=3), Qatar (N=2), Italy (N=2), Sweden (N=2),
Belgium (N=1), Bulgaria (N=1), Ecuador (N=1), Arab Emirates (N=1), France (N=1) and Mexico (N=1) (Figure 2).

Figure 2: Publications about AI in journalism, by country, in the period 2008-2019.

Regarding the university or research centres, a total of 143 institutions have been detected, based on the affiliation of the first author. The most prolific corporations are the North East University (USA), with 6 publications and the Dublin University (Ireland), with another 6 publications. There follow the American Universities of Texas with 5 scientific texts (3 in Austin and 2 in Arlington); Minnesota, with 5; and Maryland, also with 5 publications. There outstanding the Universities of Oregon (USA), Duke (USA) Hamburg (Germany) and Ljubljana (Slovenia) with 4 publications, while the Universities of Seville (Spain), Zurich (Switzerland), Florida (USA), the Herzliya Interdisciplinary Centre (Israel) and the Technological University of Nanyang (Singapore) gather 3 scientific texts.

3.3. Mastheads with the highest number of publications and authors

Regarding the publications where scientific texts are published, 99 different titles were found, being journals, the medium representing a higher percentage: 44% (N=92). Specifically, Digital Journalism (N=36), has been the journal with the greatest number of texts published. Secondly, there is Lecture Notes in Computer Science, with 16 scientific articles, followed by Journalism Practice, with 14 articles and Doxa Comunicación, with 8 articles. Likewise, there outstanding journals like Journalism (N=4), New Media & Society (N=4), Theoretical and Practical Issues of Journalism (N=4) and Journalism Studies (N=3). Among the journals, there is also the Spanish journal El profesional de la información, with 3 publications. Therefore, 9 mastheads gather almost 50% of the production related to the researched subjects.

About authorship, the trend is the presence of a single signer (N=73): practically 35% of the documents has a single singer. Regarding co-authorship, 57 publications are signed by two authors (27.3%) and 44 by three authors (21.1%). There are also researches with four (N=17), five (N=9), six (N=2) and seven authors (N=2). Likewise, there was one publication signed by eight authors, three texts with the signature of nine authors and one with twelve authors.
3.4. Research approaches

The content analysis has allowed to determine that 48.8% of researches have a descriptive approach. A significant number of them are focused on the description of AI related tools used in journalism, methods to develop specific activities (filter sources, locate verifiable facts...) or work methodologies in editorial offices.

About 39.7% of texts present an exploratory approach, while experimental orientation occurs in 5.3% of cases. Lastly, it was determined that 4.3% of the sample has an explanatory approach and 1% correlational. In another 1% of the 209 cases analysed it was not possible to determine the type of research because they were books or introductions.

3.5. Themes and keywords

In terms of themes, the analysis was focused on the study of keywords used in the English versions of scientific texts. In those cases where the article was only found in a different language, it was translated into English.

As an initial nuance, 20% of the analysed elements (N=41) did not contain keywords, most of them were conference papers (50%), chapters and books (26.2%) and journal articles (16.7%). In the remaining analysis units, a total of 1,110 keywords was detected. After screening and reviewing to eliminate duplicates, errors, and incongruences between singular and plural forms, 511 unique terms resulted thereof. The eleven most usual terms were: 1) “Computational journalism”, present in 33% of scientific texts; 2) “Automated journalism”, with 15.8%; 3) “Robot journalism”, found in 15.3% of works; 4) “Journalism”, appearing in 14.4% of production; 5) “Artificial intelligence”, present in 13.9%; 6) “Data journalism”, in 12.4%; 7) “Algorithms”, used in 12%; 8) “Automation”, found in 7.7% of researches; 9) “Algorithmic journalism”, represented in 5.7%; 10) “Social media”, found in 5.7% of the production and 11) “Big data”, present in 5.3%. Apart from this list, there are other concepts such as “natural language generation” (4.3%), “bots” (3.8%), “fact checking” (3.3%), “innovation” (3.3%), “technology” (3.3%) or “computer assisted reporting” (2.9%).

Following the methodology explained in the epigraph 2.4, the analysis allowed us to determine that the main themes analysed are related to the production of journalistic pieces (12%), data journalism (7.66%), big data with different applications –data-driven content, for example– (6.2%), the theoretical approaches (5.7%), social networks (5.7%) and information verification (5.3%). To a lesser extent, there appear themes linked to specific technological developments (2.9%), personalisation of contents (2.9%), audiences (2.9%), credibility of news generated through algorithms (2.4%), ethics (1.9%), identification of events on online media (1.9%) or the management of comments (1.9%).

3.6. Reach of academic studies

Databases consulted have offered data about the citation of 176 works. In total, 2,319 citations were reached. Five texts (2.8%) exceeded one hundred citations at the moment of study, representing 25.1% of the total. All those correspond to Universities in the USA. In the two most cited articles, there is the author Nick Diakopoulos, who is also included in the fifth.

Between 99 and 76 citations position in 1.7% of texts, gathering 11% of citations. About 3.4% have been cited between 50 and 75 times (16.4% from the total of citations), while about 6.2% hoard between 26 and 50 references (18.2% from the total of citations). Most texts (63.1%) have been cited less than 25 times. This group accumulates 29.19% of all citations. Lastly, 22.7% of the documents with available data, did not receive any citations.
Table 3: Most cited articles about AI and journalism.

<table>
<thead>
<tr>
<th>Title</th>
<th>Author</th>
<th>Year</th>
<th>Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond in the rough: Social media visual analytics for journalism inquiry</td>
<td>Diakopoulos, N., Naaman, M. and Kirvan-Swaine, F.</td>
<td>2010</td>
<td>139</td>
</tr>
<tr>
<td>Algorithmic accountability. Journalism investigation of computational power structures</td>
<td>Diakopoulos, N.</td>
<td>2015</td>
<td>122</td>
</tr>
<tr>
<td>Clarifying journalism’s quantitative turn. A typology for evaluating data journalism, computational journalism and computer-assisted reporting</td>
<td>Coddington, M.</td>
<td>2015</td>
<td>112</td>
</tr>
<tr>
<td>Towards a sociology of computational and algorithmic journalism</td>
<td>Anderson, C.W.</td>
<td>2013</td>
<td>106</td>
</tr>
<tr>
<td>Finding and assessing social media information sources in the context of journalism</td>
<td>Diakopoulos, N De Choudhury, M., Naaman, M.</td>
<td>2012</td>
<td>104</td>
</tr>
</tbody>
</table>

Source: Own elaboration based on the analysis of WOS and Scopus.

4. Conclusions

This research meets the objective established initially in terms of analysing recent studies (2008–2019) about the application of AI in the journalistic field. The two databases analysed, Web of Science and Scopus, show a close coincidence in the resulting publications. From the analysis of the 209 scientific publications, we observe that there is a terminological variety for referring to the elaboration process and the distribution of news elaborated and published, based on the procedures carried out by machines where there is no intervention of any journalist whatsoever to write or publish the text. Some concepts like robot journalism, algorithm journalism, automated journalism, computational journalism, augmented journalism, artificial journalism, high tech journalism or AI are used for the same purpose. Therefore, the proposals of Vállez and Codina (2018) are confirmed, who indicate that the terms used to call this discipline is quite varied because there is no clear line limiting its reach.

It is observed that the first texts included in WOS and Scopus were published in 2010, but it was the past five years when a gradual increase is produced. An example of this is that in 2015, 18 publications are detected, while in 2019 the number increases up to 61. This fact can be justified due to the relative novelty of this discipline and the lack of existence of specific research groups, besides a scientific context that is not so prone to this sort of studies, because in the Spanish context, for instance, the communication research is mainly oriented to the study of journalistic contents (Martínez Nicolás & Saperas, 2011). Therefore, the first question presented in this research is answered.

In the same line, the results allow to respond to the second and the third research question, when indicating that the main scientific production comes from the North American scientific field. Among the countries with the greatest number of studies conducted in this area, the United States stand out (with 69 publications, a number that represents 33% of the total). Together with the increase of the volume of publications, the development of joint projects among countries and/or universities is appropriate, in order to analyse in a richer and more complete manner, a phenomenon with a clearly global tenor.

Journals (92) are the ones concentrating the greatest number of scientific publications, therefore there is a certain specialisation or trend of authors to disseminate their findings in this sort of publications. In this sense, Digital Journalism, Lecture Notes in Computer Science and Journalism Practice stand out, concentrating 31% of publications. Likewise, the Spanish journal Doxa Comunicación positions among the journals with greatest number of researches, with 8 articles. This is due to the fact that the masthead published, in May 2019, a monographic article about AI and Journalism, coordinated by the professors José Miguel Túñez López, from
the University Santiago de Compostela, and Santiago Tejedor Calvo, from Universidad Autónoma de Barcelona. This trend confirms the theories of Krauskopf and Vera (1995), Vesuri (1995) and van Raan (2001) about the scientific journals being the means ‘par excellence’ of the scientific communication. But evidently, the article is not just the only format used by researchers to disseminate the results of their work, and coexists with other scientific products, like books, book chapters, conference papers, introductions, and reviews, among others, that are being outcast to a subsidiary position in the ecosystem of the scientific publication (RQ4).

Regarding the characteristics of the authorship, the presence of a single author prevails, something that occurs practically in 35% of publications reviewed. The presence of two signers is also noteworthy, with 27.3%, hence answering the fourth question of this research. These data are aligned with the works of Baladrón, Machado and Correyero (2017) and Martínez Nicolás and Carrasco Campos (2018), who evidence the trend of works with a single authorship and that underline the slow, but sustained growth, of co-authorships (RQ5).

Considering that we are facing a relative new field, it is logical that a good part of researches have an exploratory (39.7%) and descriptive approach of tools, methodologies, and applications (48.8%) (RQ6).

Regarding themes, the greatest relevance of the research about the robot journalism has focused on the informative production, data journalism, management of large databases in order to, among other questions, orientate the contents towards the interests of the audience, the theoretic exploration, the application in the context of social networks and the information check (RQ7). The transversality of this practice allows to open a new field of knowledge that integrates different disciplines and entails a new form of interacting with information and a new form of diffusion in media. In this line, Nicolescu (2002) reminds that the boundaries of individual disciplines are transcended, and multiple perspectives are used to generate emergent knowledge.

This work, despite answering the seven research questions, has certain limitations and leaves possibilities open for future researches. The fact of circumscribing analysis to the work with international visibility published in Web of Science and Scopus, excludes other type of research results. Therefore, there is great room for the development of similar reviews, that could provide further knowledge about the thematic trends and approaches used and could serve to add value to the research and academic centres bearing the banner of robot journalism. Likewise, there are pending matters, like the analysis of interdisciplinary works involving researchers from different areas or fields of great interest, for the theoretical and practical approach of the object of study, like computer science, documentation, statistics and linguistics.

Luis Mauricio Calvo Rubio receives funding from the European Regional Development Fund (ERDF) call 2020/377. This article is funded by the COMPUBLIC research group, within the framework of the research activities of the Own Plan of the University of Castilla-La Mancha. It is also part of the actions of the research project titled “News, networks and users in the hybrid media system. Transformation of the media and news industry in the post-industrial era” (RTI2018-095775-B-C43) (Mineco/Feder), Ministry of Science, Innovation and Competitiveness of Spain (2019-2021).

References


