



Article

A Methodological Proposal to Evaluate Journalism Texts Created for Depopulated Areas Using AI

Luis Mauricio Calvo Rubio ^{*}, María José Ufarte Ruiz and Francisco José Murcia Verdú

Faculty of Communication, Universidad de Castilla-La Mancha, 16071 Cuenca, Spain; mariajose.ufarte@uclm.es (M.J.U.R.); francisco.murcia@uclm.es (F.J.M.V.)

^{*} Correspondence: luismauricio.calvo@uclm.es

Abstract: The public service media Radio Televisión Española (RTVE) conducted a proof-of-concept study to automatically generate reports on the results of the local elections of 28 May 2023 in Spanish communities with fewer than 1000 inhabitants. This study describes the creation, testing and application of the methodological tool used to evaluate the quality of the reports generated using artificial intelligence in order to optimize the algorithm. The application of the proposed datasheet provided a systematic analysis, and the iterative use of the tool made it possible to gradually improve the results produced by the system until a suitable threshold was reached for publication. The study also showed that, despite the ability of AI systems to automatically generate a large volume of information, both human labour and the reliability of the data that feed the system are essential to ensure journalistic quality.

Keywords: artificial intelligence; journalism; public media; automated journalism; information deserts; journalistic quality



Citation: Calvo Rubio, Luis Mauricio, María José Ufarte Ruiz, and Francisco José Murcia Verdú. 2024. A Methodological Proposal to Evaluate Journalism Texts Created for Depopulated Areas Using AI. *Journalism and Media* 5: 671–687. <https://doi.org/10.3390/journalmedia5020044>

Academic Editor: Andreu Casero-Ripollés

Received: 1 April 2024
Revised: 11 May 2024
Accepted: 18 May 2024
Published: 27 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The arrival of artificial intelligence (AI) systems in newsrooms has made it possible, amongst other things, to generate automated news content (Tejedor 2023). Although the algorithms were initially used to produce simple pieces based on structured data related to sports, finance or the weather, over time, initiatives have emerged that have enabled the creation of more elaborate texts that, in addition to relieving journalists of less creative work, have provided the means to fill some information gaps that would not otherwise be possible to fill, due to the limited resources of media companies, as noted by Aramburu Moncada et al. (2023). One such case concerns election information in sparsely populated towns.

With this aim in mind, in late 2023, the Editorial Board for Technology, Innovation and Systems of Radio Televisión Española (RTVE) conducted a proof-of-concept study to produce reports on the results of the local elections held in Spain on 28 May 2023. A multidisciplinary team made up of engineers, journalists and information technology experts from public institutions and private enterprises worked to create a system based on AI that could automatically generate reports related to the election results in 4941 Spanish communities with fewer than 1000 inhabitants.

From the time the polling stations closed at 10 p.m., the system generated 59,052 pieces with text, images and graphics as the results came in. The information, based on official data from the Ministry of the Interior, could be consulted on the website www.rtveia.es. Additionally, 332 h of audio using a synthetic voice were generated and also accessible on the same page through the virtual assistant Alexa. Because of the good results, the system was used again in the general elections of 23 June 2023. On that occasion, 23,006 reports were generated on the evolution of the turnout and 53,034 on the results of the vote. The significance of the project was confirmed when it was recognized with a prestigious international IBC2023 award in the category of Social Impact, which acknowledges groundbreaking initiatives that are redefining the landscape of the media industry.

To train the system, it was necessary to create a methodological tool that could calibrate the quality of the texts being generated by the machine and propose any required improvements. This study examines the undertaking based on the experience of a team of researchers from the Universidad de Castilla-La Mancha, who worked alongside RTVE journalists and technicians from Narrativa, the first international automated news agency.

2. AI and the Generation of Texts for Election Reports

2.1. Automated Texts and Their Impact in Rural Areas

In the media ecosystem today, the use of new technologies based on artificial intelligence for election coverage has been accompanied by changes in working methods and dynamics, reducing costs and automating essential work that had, until recently, been considered to belong exclusively to the realm of humans in the journalism profession (Brennen et al. 2022). Although a number of works have approached this topic, few extensive field studies have been conducted. The impact of this tool on the generation of election news has been analysed by Fanta (2017), Sánchez Gonzales and Sánchez González (2017) and Digiday (2017), amongst others. For Diaz-Garcia et al. (2020), its function ranges from providing and comparing information on the candidates and their proposals to compiling data on the preferences of the electorate in order to formulate party strategies, campaign messages and other political communication material, to predicting non-official election outcome scenarios, amongst other applications. In short, AI helps journalists to generate new news content by searching for and analysing data that are personalized and adapted to the needs of different audiences who did not previously receive this information due to a lack of resources (Canavilhas 2015).

However, Aramburu Moncada et al. (2023) argue that a key potential aspect of this technological tool lies in interpreting the election results from small towns in depopulated areas of Spain and transforming this information into automated reports without human intervention. They highlight the initiatives undertaken in this country in 2019 by the digital newspaper *El Confidencial* and the start-up Narrativa, the agency that went on to collaborate with the public broadcast entity RTVE in 2023 to automatically generate articles and audios about the election outcomes and developments in Spanish communities with fewer than 1000 inhabitants. The resulting accurate and intelligible synthetic words and texts once again demonstrated the scope and importance of artificial intelligence with regard to extending into areas that traditional coverage cannot reach, as proposed by LeCompte (2015).

2.2. Generative AI Tools and Models

The appearance of artificial intelligence models and tools has radically transformed the way the profession of journalism is understood and performed, changing the education process, the production of content and the knowledge and skills required by professionals. The arrival of AI has led academics to raise various questions about the efficacy and limitations of these applications, especially ChatGPT, a natural language processing model (NLP) (Guida and Mauri 1986) developed by the company OpenAI (OpenAI 2022).

A number of works have analysed the positive and negative sides of AI, including Aydin and Karaarslan (2022), Lopezosa (2023) and Wang et al. (2023), amongst others, who highlight the efficiency of generative artificial intelligence in the production of content but focus on the need to be responsible and ethical, since this tool does not disclose the sources of information or body texts it uses to search for information and create responses. Other studies, such as the work by Lopezosa et al. (2023), have expanded their scope to new tools like Midjourney, Dall-e and Stable Diffusion, recognizing that, although they are capable of reaching the general public, they make mistakes in their application and collection of data. Consequently, the authors recommend that humans provide feedback when they are used, in order to reinforce all the verification procedures.

Turning to image and video creation, Guerrero-Solé and Ballester (2023) argue that the applications Stable Diffusion, Midjourney and Dall-e have changed the paradigm, although they call for additional studies and research to better understand the opportunities,

limitations and related challenges. In a study of the audiovisual creations generated by GenAI, [López Delacruz \(2023\)](#) observed that one of its primary limitations is that it imitates earlier visual styles without contributing any narrative innovation.

Faced with this situation, [Van-Dis et al. \(2023\)](#) recommend that both researchers and journalists apply a critical approach to the use of artificial intelligence tools, carefully verifying the results, the data and the references, since these tools contain biases resulting from the models used or their training data.

2.3. Previous Studies

Academic studies on the quality of news articles created by IA have increased and improved in recent years ([Jung et al. 2017](#); [Waddell 2019](#); [Tandoc et al. 2020](#); [Jia and Johnson 2021](#); [Wölker and Powell 2018](#); [Lermann Henestrosa et al. 2023](#)). The first international research studies, which appeared in the mid-2000s, primarily focused on how these tools could write texts by themselves ([Carlson 2015](#)) and on how articles written by robots are perceived, using indicators to measure their quality ([Clerwall 2014](#); [Haim and Graefe 2017](#); [Zheng et al. 2018](#)).

In Spain, [Ufarte Ruiz and Murcia Verdú \(2018\)](#) were amongst the first to analyse the quality of the information in political and financial reports produced by Gabriele, the Narrativa software, using a questionnaire conducted with more than 100 journalists. [Calvo-Rubio and Ufarte-Ruiz \(2020\)](#), in turn, surveyed almost 200 Journalism and Audiovisual Communication students in a number of Spanish universities and concluded that the quality of automated news is deficient due to the lack of contrast, absence of interpretation, non-existence of humanity and sensitivity and poor wording.

[Rojas Torrijos and Toural-Bran \(2019\)](#) shifted the focus to the sports coverage produced by AnaFut, the bot developed by the digital paper *El Confidencial*, to identify the statistics that facilitate orderly data handling and the programming of routine news productions, given the cyclical and repetitive nature of matches and tournaments. In a similar vein, [Murcia Verdú et al. \(2022\)](#) analysed the content of 28 news stories to discover whether these types of texts have the same quality standards as pieces written by journalists. On the other hand, [Rojas Torrijos \(2021\)](#) and [Calvo-Rubio and Rojas-Torrijos \(2024\)](#) have addressed the importance of reinforcing the supervision of journalistic ethics in semi-automated journalism.

In general, the quality of automated news is perceived as excellent, although with some limitations, like the impossibility of adding context, different points of view and interpretation ([Sandoval-Martín and Barrolleta 2023](#)).

In the field of public service media (PSM), the demands of quality and compliance with the ethical principles of journalism reach the highest level due to their relevance, need for trust and social function. In this field, [Fieiras-Ceide et al. \(2023\)](#) have studied the use of AI in the recommender systems of 14 European public broadcasters. In a more general scope, [Zaragoza and García Avilés \(2022\)](#) and [Direito-Rebollal and Donders \(2023\)](#) have focused on analysing how PSMs are adapting to the new communicative context where technological innovation plays an essential role, with special emphasis on quality and content adaptation.

2.4. The RTVE Project

As [Fieiras-Ceide et al. \(2023, p. 354\)](#) explain, “the renewed digital context motivates public broadcasting corporations to internalize innovative processes that allow them to be relevant in people’s lives. These processes are not only limited to the development and integration of sophisticated technological prototypes, but are closely linked to a philosophy of constant change and renewal of ideas and ways of thinking”. In the case of Radio Televisión Española (RTVE), the importance of Artificial Intelligence as a fundamental tool for the creation, production and distribution of content has been emphasized.

Since 2021, the Editorial Board for Technology, Innovation and Systems of RTVE has been working to create a tool that would make it possible to generate news on election

results in communities with fewer than 1000 inhabitants, in order to ‘offer a service that is not possible to provide using traditional media’ to citizens and ‘assist RTVE professionals in their jobs by creating an initial version on which they can work’ (RTVE 2023). The aim of these initiatives on the part of the public broadcasting service is to emphasize its mission of public service. Earlier, in 2019, the BBC News Lab had already experimented with the automated generation of news related to general elections using the engine called SALCO (Semi-Automated Local Content). Subsequently, in May 2023, they developed a project to generate hyper-local news based on official statistical data (Hatcher-Moore 2023).

For the necessary technological developments, RTVE used the services provided by the company Narrativa, specifically its natural language generation system known as Gabriel. The company trained the software to write election reports using texts written by journalists as the corpus. The first tests were conducted during the elections for the Madrid Assembly in 2021. At that time, 2600 pieces were generated in four hours that were not posted but were used to adjust the system (Aramburu Moncada et al. 2023). After that trial, attention turned to the local elections of 2023. The challenge was to adjust the system to offer useful, real-time information on the vote count in Spanish communities with fewer than 1000 inhabitants. RTVE journalists designed the structure of the reports that the system needed to generate using the official structured data supplied by the Ministry of the Interior throughout election day as the source.¹

To train the system, data from earlier local elections were loaded and the phase of testing and adjusting the algorithm began. This work continued throughout the first four months of 2023. The project also participated with the studio Monoceros Labs and the Universidad de Granada to create the synthetic voice, with the technological support of Amazon Web Services (AWS). The Spanish National Organization of the Blind, ONCE, provided advice about information accessibility.

In order to evaluate the quality of the automatically generated reports and measure the advances, it was necessary to create a methodological tool. This task was assigned to the Universidad de Castilla-La Mancha, thus expanding the university’s collaboration with RTVE’s territorial hub in the region of Castilla-La Mancha, a model of technological innovation for the public broadcasting service.

3. Methods

The proposed methodology uses content analysis; although mainly quantitative, it is complemented with value judgements in the form of observations to facilitate the implementation of improvements, in line with that carried out by Odrizola-Chéné et al. (2020).

The design of the methodology began with the creation of a datasheet containing 11 variables and 58 dimensions that could be used to detect anomalies and provide a numerical evaluation of the final result (Table 1). Six experts in the fields of journalistic writing and cyber journalism were involved in the design and evaluation to ensure the reliability of the instrument.

Table 1. Datasheet.

Characteristic	Cluster	Variable	Dimension
Journalistic quality	Headline elements	Billboard	Autonomy
			Clarity
			Concision
			Spellcheck
			Grammar check
	Headline	Headline	Informative
			Dynamic
			What/Who

Table 1. Cont.

Characteristic	Cluster	Variable	Dimension
Journalistic quality	Headline elements	Headline	Summarizes the news
			Brevity
			Concision
			Clarity
			Complete structure
			Precision
			Autonomy
			Absence of punctuation marks
			Spellcheck
			Grammar check
	Callouts	Callouts	Typographical difference
			Autonomy
			Information relevance
			Precision
			Spellcheck
			Grammar check
			Highlighted elements
	Text	First paragraph	No summary
			Begins without adverb or adverbial
			Spellcheck
			Grammar check
			Context
		Body text	Background
			Facts
			Informative interpretation of data
			Spellcheck
			Grammar check
		Overall assessment	Clarity
			Concision
			Coherence
			Cohesion
Suitable for the medium used	Online media elements	Links	Presence
			Relevance
			Accurate anchor text
		Photos	Presence
			Relevance
			Caption (present and correct)
		Graphics	Presence
			Relevance
			Headline
		Audio	Presence
			Relevance
			Headline

Table 1. Cont.

Characteristic	Cluster	Variable	Dimension
Suitable for public medium	Complies with RTVE policies		Public relevance of the information
			Accuracy
			Objectivity
			Impartiality
			Discriminatory language
			Inclusive language
			Precise use of data

Source: Authors.

This tool was iteratively applied to a representative sample of the total reports generated by the system with the data loaded for the training. After each analysis, the system was adjusted before it generated new reports. The process was repeated until the result was accepted as valid by the multidisciplinary team created to develop the project led by RTVE.

The datasheet was configured on the basis of three characteristics: the journalistic quality of the text, its suitability for the medium used for transmission and its compliance with the policies of a public information service.

For the first characteristic, the textual elements of the reports were identified and, based on the academic literature (Trillo-Domínguez and Alberich-Pascual 2017; López Hidalgo 2001, 2019; Armentia Vizuete and Marcet 2003), the characteristics that needed to be accounted for were determined. As a result, the dichotomous variables were obtained that evaluated the presence, or lack thereof, of each of the characteristics (0: not present; 1: present), to which four variables were added to evaluate the overall clarity, concision, coherence and cohesion of the final result (Grijelmo 2014; Martínez Albertos 1974; Dovifat 1959).

For the second characteristic, the team evaluated the suitable appearance of the narrative elements characteristic of a webpage, the main platform for posting the content: links, photos, audios and graphics. Videos were omitted as they were not included in the proposed news format. All the elements were studied for their presence, or lack thereof, and, in the former case, whether they were relevant in terms of providing valuable information. The suitability of the anchor texts for the links, the use of photo captions and the presence of a descriptive headline for the graphics were also reviewed (Salaverria 2005; Cobo 2012).

Finally, a series of variables related to the suitability of the content for a public medium were included: the public relevance of the information, the accuracy, objectivity, impartiality, use of nondiscriminatory language, use of inclusive language and the precise use of the data (Mandato-Marco a la Corporación RTVE 2008; Estatuto de Información de la Corporación RTVE 2008; Manual de Estilo de RTVE 2010; Guía de Igualdad de RTVE 2022).

To test the tool and the reliability of the data obtained from the six encoders—experts involved in its development—a pilot test was conducted based on analysing 12 of the 632 automatically generated reports (1.9% of the total). Krippendorff's (2004) alpha was employed for the study index, yielding an average result of 0.681, divided by category as follows (Table 2):

Table 2. Inter-rater reliability analysis using Krippendorff's alpha.

Headlines	Text	Internet Resource	Public Medium	Medium
0.424	0.549	0.900	0.850	0.681

Source: Authors.

As the main discrepancies were detected in the evaluation of the headlines and the body text, the coding was revised and the analytical criteria unified in order to raise the index above 0.7 in all the categories, a figure considered sufficient to obtain reliable data (Hayes and Krippendorff 2007).

4. Results

After adjusting the datasheet and the codebook, the tool was applied to 106 reports generated by the system under study to a number of communities. The results (Table 3) identified the elements that required optimization.

Table 3. Results of the first sample.

Cluster	Variable	Dimension	% Reports That Comply
Headline elements	Billboard	Autonomy	49.06%
		Clarity	100%
		Concision	100%
		Spellcheck	100%
		Grammar check	100%
	Headline	Informative	98.11%
		Dynamic	100%
		What/Who	100%
		Summarizes the news	97.17%
		Brevity	100%
		Concision	97.17%
		Clarity	97.17%
		Complete structure	100%
		Precision	92.45%
		Autonomy	100%
		Absence of punctuation marks	100%
		Spellcheck	100%
		Grammar check	99.06%
	Callouts	Typographical difference	83.96%
		Autonomy	83.96%
		Information relevance	100%
		Precision	93.4%
		Spellcheck	96.23%
		Grammar check	87.74%
Text	First paragraph	Highlighted elements	100%
		No summary	100%
		Begins without adverb or adverbial	0%
		Spellcheck	89.62%
		Grammar check	95.28%
		Context	82.08%
	Body text	Background	99.05%
		Facts	98.11%
		Informative interpretation of data	52.83%
		Spellcheck	57.55%
		Grammar check	83.02%
		Clarity	2.13
	Overall assessment	Concision	2.06
		Coherence	1.86
		Cohesion	1.73

Table 3. Cont.

Cluster	Variable	Dimension	% Reports That Comply
Online media elements	Links	Presence	100%
		Relevance	100%
		Accurate anchor text	100%
	Photos	Presence	100%
		Relevance	27.36%
		Caption (present and correct)	0%
	Audio	Presence	100%
		Relevance	100%
		Headline	16.04%
	Graphics	Presence	99.06%
		Relevance	99.06%
		Headline	50.94%
Suitable for public medium		Public relevance of the information	99.06%
		Accuracy	90.57%
		Objectivity	93.4%
		Impartiality	94.34%
		Discriminatory language	100%
		Inclusive language	0%
		Precise use of data	87.74%

Source: Authors.

In the section ‘Headline elements’, it was confirmed that all the billboards were clear, concise and accurate from the perspective of spelling and grammar. However, several encoders concluded that they were not autonomous but required other elements to understand them accurately. Almost all the headlines were informative (98.11%), dynamic (100%), answered the questions ‘what’ and ‘who’ (100%), summarized the news (97.17%), were brief (100%), were concise and clear (97.17%), were autonomous (100%), did not use punctuation marks (100%) and used correct spelling (100%) and grammar (99.06%). Of the pieces, 92.45% were also sufficiently precise. Finally, all the callouts were relevant (100%). The other characteristics were present to varying degrees: autonomy (83.96%), typographical difference (83.96%), precision (93.4%), spellcheck (96.23%) and grammar check (95.28%).

The analysis of the body text showed that the first paragraphs reflected the most important elements of the reports (100%) but without summarizing (100%). However, all of them began with an adverb (100%). In 89.62% of the cases, the spelling was correct and 95.28% of the reports followed the rules of grammar. The articles generated by AI included a background (99.05%), provided context data (82.62%) and clearly explained the facts (98.11%). However, anomalies were detected in the interpretation of the data in 47.2% of the pieces, as well as in the spelling (42.45%) and grammar (16.98%). The overall evaluation received 2.13 out of 3 points for clarity, 2.06 for concision, 1.86 for coherence and 1.73 for cohesion.

In the section related to the elements of digital media that contribute to the story, the links were present and relevant and the anchor text was suitable in 100% of the cases. All the pieces also included photos, but it was concluded that they were only relevant in 27.36% of the reports, since many consisted of mere resource graphics that did not contribute any information. In no cases were captions present. Audios (100%) and graphics (99.06%) were also present, with a high degree of relevance (100% and 99.06%, respectively). However,

in nearly half of the reports (49.06%), the headline for the graphics was either not present or incorrect.

Finally, the reports were determined to have public relevance (99.06%) and be accurate (90.57%), objective (93.4%) and impartial (94.34%). Additionally, the language was not discriminatory (100%) and the data were precise (87.74%). Possible improvements in the use of inclusive language were identified in every report.

The observations included by the encoders were used to create a list of elements to improve in each of the sections, with references to the text where they had been identified (Table 4):

Table 4. List of elements to improve after sample 1.

Headline Elements	Discrepancies between the Callouts and the Text
Text	Including sources of information in the text Including the total number of councillors on the council in constructions like ‘The party obtained a total of X councillors’ out of a total of X. Replacing ‘a pact with the other parties to be able to govern’ with a phrase that does not indicate the need for a pact with all of them. Data interpretation: discrepancies between the headline, callouts and information (in the case of an absolute majority) In communities with an absolute majority, repeating this circumstance at several points, making the text seem repetitive and less coherent Starting paragraphs with adverbials Incorrect concordances Missing words Discrepancies between callouts and text Redundancies Misinterpreting data with regard to pacts to form a government Presenting a fair victory
Online media elements	Need to include photo and author captions Need to include photos with informative content Need to include captions and headlines with graphics
Suitability for public medium	Using generic terms (for example, ‘socialists’ for the Socialist Party or ‘inhabitants’ or ‘residents’ for the electorate or citizens)

4.1. Second Analysis

After adjustments made by the narrative technical team based on the results obtained, 633 new reports were generated, of which 553 received a valid analysis. Since the billboards, which the team considered accurate, did not vary between the reports, they were eliminated from this evaluation.

The new data were compared with those obtained in the first analysis below in Table 5.

The percentage of reports that complied with the characteristics established for the headline elements improved for almost all the items analysed. The only notable decrease was found in the typographical difference in the billboards. The reason for this was a discrepancy between encoders; while some understood that these texts required larger letters than the body text, others understood that the full stop that separated them from the beginning of the sentence was sufficient. In the end, the presentation was accepted as valid, as it adhered to the style used by RTVE on its website.

Table 5. Results of the second sample.

	Element	Characteristic	% Reports That Comply	Results from First Sample
Headline elements	Billboard	Autonomy	Not evaluated	49.06%
		Clarity		100%
		Concision		100%
		Spellcheck		100%
		Grammar check		100%
	Headline	Informative	100%	98.11%
		Dynamic	100%	100%
		What/Who	99.85%	100%
		Summarizes the news	100%	97.17%
		Brevity	100%	100%
		Concision	99.54%	97.17%
		Clarity	99.21%	97.17%
		Complete structure	100%	100%
		Precision	98.75%	92.45%
		Autonomy	100%	100%
		Absence of punctuation marks	100%	100%
		Spellcheck	99.84%	100%
		Grammar check	100%	99.06%
	Callouts	Typographical difference	66.51%	83.96%
		Autonomy	88.31%	83.96%
		Information relevance	99.1%	100%
		Precision	97.31%	93.4%
		Spellcheck	91.47%	96.23%
Text	First paragraph/lead	Grammar check	88.63%	87.74%
		Highlighted elements	99.68%	100%
		No summary	99.84%	100%
		Begins without adverb or adverbial	0.32%	0%
		Spellcheck	91%	89.62%
	Body text	Grammar check	99.68%	95.28%
		Context	83.25%	82.08%
		Background	100%	99.05%
		Facts	99.21%	98.11%
		Informative interpretation of data	57.5%	52.83%
		Spellcheck	44.85%	57.55%
		Grammar check	98.92%	83.02%
	Overall assessment	Clarity	1.84	2.13
		Concision	2.05	2.06
		Coherence	2	1.86
		Cohesion	1.86	1.73

Table 5. Cont.

Element	Characteristic	% Reports That Comply	Results from First Sample
Online media elements	Links	Presence	100%
		Relevance	100%
		Accurate anchor text	100%
	Photos	Presence	100%
		Relevance	49.37%
		Caption (present and correct)	38.88%
	Audio	Presence	100%
		Relevance	100%
		Headline	0.1%
	Graphics	Presence	100%
		Relevance	97.29%
		Headline	100%
Suitable for public medium	Public relevance of the information	99.82%	99.06%
	Accuracy	95.12%	90.57%
	Objectivity	98.01%	93.4%
	Impartiality	97.29%	94.34%
	Discriminatory language	99.84%	100%
	Inclusive language	2.37%	0%
	Precise use of data	80.73%	87.74%

Source: Authors.

The variables related to the text also showed improvement. However, there was a notable decline in the number of articles that met the spellcheck criteria (44.85%). A review of the observations included by the encoders on the datasheets determined that the problem lay in the data used to generate the graphics; these texts did not include the accent marks required in the Spanish language.

The principal problems were identified with the precision and the interpretation of the data. Once again, the errors were largely related to the graphics; the graphic representation did not coincide with the text content; the percentages did not add up to 100%; and there were discrepancies between the number of votes and the census, amongst other problems. The result was a decrease in the average score for clarity (1.84).

RTVE made efforts to obtain at least one photograph from each community. The decision was also made to include another image related to election processes from their documentary archives, which increased the relevance of the images (49.37%). Photo captions also began to appear (38.88%).

All the parameters related to the suitability of the texts for a public medium improved, but the use of inclusive language only reached 2.47%. A detailed analysis of this section revealed that the reports complied with RTVE policies, although there was room for improvement in more general terms.

The proposals for improvement centre around three elements: (1) the problems detected in the graphics data; (2) the interpretation of the data in the text (incoherence, lack of clarity, contradictions, declaration of a winning party in some cases where a pact was necessary); and (3) style corrections (inaccurate or missing conjunctions, confusing wording, incorrect expressions).

4.2. Third Analysis

During the first week of April, following the introduction of the changes suggested by the technicians, the system generated a further 633 items, of which 565 received a valid analysis. The results are presented in Table 6.

After correcting the deficiencies detected in each of the reports, the working team validated the suitability of the content for final publication using the data provided by the Ministry of the Interior on election day. The results can be consulted at <https://www.rtve.es/elecciones-municipales-2023> (accessed on 20 March 2024).

Table 6. Results of the third sample.

	Element	Characteristic	% Articles That Comply	Results from Second Sample	Results from First Sample
Headline elements	Billboard	Autonomy	Not evaluated	Not evaluated	49.06%
		Clarity			100%
		Concision			100%
		Spellcheck			100%
		Grammar check			100%
	Headline	Informative	100%	100%	98.11%
		Dynamic	100%	100%	100%
		What/Who	100%	99.85%	100%
		Summarizes the news	100%	100%	97.17%
		Brevity	100%	100%	100%
		Concision	99.47%	99.54%	97.17%
		Clarity	99.47%	99.21%	97.17%
		Complete structure	100%	100%	100%
		Precision	99.65%	98.75%	92.45%
		Autonomy	100%	100%	100%
		Absence of punctuation marks	100%	100%	100%
		Spellcheck	99.82%	99.84%	100%
		Grammar check	99.65%	100%	99.06%
	Callouts	Typographical difference	Not evaluated	66.51%	83.96%
		Autonomy	80.35%	88.31%	83.96%
		Information relevance	99.82%	99.1%	100%
		Precision	99.29%	97.31%	93.4%
		Spellcheck	87.79%	91.47%	96.23%
		Grammar check	81.42%	88.63%	87.74%

Table 6. Cont.

	Element	Characteristic	% Articles That Comply	Results from Second Sample	Results from First Sample
Text	First paragraph/lead	Highlighted elements	100%	99.68%	100%
		No summary	99.84%	99.84%	100%
		Begins without adverb or adverbial	0.35%	0.32%	0%
		Spellcheck	87.96%	91%	89.62%
		Grammar check	95.75%	99.68%	95.28%
	Body text	Context	82.65%	83.25%	82.08%
		Background	99.47%	100%	99.05%
		Facts	99.82%	99.21%	98.11%
		Informative interpretation of data	62.30%	57.5%	52.83%
		Spellcheck	45.49%	44.85%	57.55%
		Grammar check	87.26%	98.92%	83.02%
	Overall assessment	Clarity	2.02	1.86	2.13
		Concision	2.27	2.04	2.06
		Coherence	2.08	2	1.86
		Cohesion	1.91	1.88	1.73
Online media elements	Links	Presence	100%	100%	100%
		Relevance	100%	100%	100%
		Accurate anchor text	100%	100%	100%
	Photos	Presence	99.82%	100%	100%
		Relevance	82.83%	49.37%	27.36%
		Caption (present and correct)	96.28%	38.88%	0%
	Audio	Presence	100%	100%	100%
		Relevance	100%	100%	100%
		Headline	43.38%	0.1%	16.04%
	Graphics	Presence	100%	100%	99.06%
		Relevance	98.58%	97.29%	99.06%
		Headline	100%	100%	50.94%
Suitable for public medium		Public relevance of the information	100%	99.82%	99.06%
		Accuracy	98.05%	95.12%	90.57%
		Objectivity	99.12%	98.01%	93.4%
		Impartiality	97.52%	97.29%	94.34%
		Discriminatory language	100%	99.84%	100%
		Inclusive language	5%	2.37%	0%
		Precise use of data	95.93%	80.73%	87.74%

Source: Authors.

5. Discussion

In view of the results, the use of the methodological tool created to evaluate the quality of the reports generated by artificial intelligence tools for the results of the local elections of 28 May 2023 in Spain improved the system. Based on the algorithm created for the elections to the Madrid Assembly in 2021, the RTVE journalists created the information structure that the AI system needed to generate, using the official election data provided by the Ministry of the Interior as its source. At that point, the tests to train the information system were carried out until a result was obtained that was considered valid for publication under the name of this public broadcasting entity.

During the training phase, a datasheet created ad hoc was used to facilitate the systematic study of the critical elements related to quality, the use of narrative mechanisms and compliance with the ethical policies of public service media. This tool made it possible to identify the weakest points, analyse them more thoroughly and make improvements when deemed necessary.

Initially, the principal problems were detected in the sections related to precision and data interpretation. The first analysis detected anomalies in the headlines (7.55%), callouts (6.6%) and body text (47.17%). The adjustments made it possible to conclude the training with a reduction in errors in these aspects, to 0.35% in the headlines, 0.71% in the callouts and 37.7% in the body text. However, the spelling and grammar checks and style were more problematic. In this case, despite the improvements to the text generation system, errors in the database used for the tests limited the extent to which the reports were able to improve in order to reach full accuracy. Nonetheless, there was a clear improvement in the texts. The overall assessment of the reports regarding concision (+0.21), coherence (+0.22) and cohesion (+0.15) increased after the successive revisions. There was almost no variation in the quality, with the final score above 2 out of 3.

The actions taken regarding the headlines, text, graphics and images improved the parameters that established the suitability of the pieces for the RTVE statute. Public relevance rose 0.94 points, reaching 100% in the end. Accuracy increased to 99.12% of the content (+7.48%), while 99.12% of the pieces were considered objective in the final sample (+5.72%), and 97.52% were determined to be impartial (+3.18). No article was found to contain discriminatory language. Although there was room to improve the use of inclusive language in a significant number of the reports generated—95% of the final sample—the results complied with the policies established by the public broadcasting entity.

The other characteristics had acceptable values from the beginning. The accuracy of the headline elements surpassed 99% for all the variables, and these numbers stayed constant or improved throughout the training. The first paragraphs highlighted the most important elements of the information, avoided summarizing and met spelling standards (−1.66%).

6. Conclusions

The proposed methodological tool enabled the assessment of the different dimensions related to quality and was useful for identifying anomalies, proposing improvements to programmers and fine-tuning algorithms. The application of the datasheet and numerical results facilitated systematic analysis, and its iterative use gradually improved the results obtained with the system to a threshold suitable for publication. With slight modifications, the datasheet can be applied to reports from other domains and means for training AI systems.

Despite the ability of AI systems to generate a large volume of automated information, human work is essential during each phase of the process, in a similar view to previous research (Van-Dis et al. 2023; Aydın and Karaarslan 2022; Lopezosa 2023; Wang et al. 2023; Lopezosa et al. 2023; Calvo-Rubio and Rojas-Torrijos 2024). The choice of the subject, the design of the information structure, the selection of the database, the system programming, the detection of anomalies during training, decisions about what improvements to make, adjusting the software and conducting the final review are all critical phases that require human participation. Moreover, aside from the steps related to programming and managing

technological systems, the journalist plays a fundamental role in this process. Be that as it may, however, the reliability of the databases used is critical when using generative AI systems.

This collaboration between AI systems and journalists in the new media ecosystem reinforces the idea that news professionals should not see technological tools as enemies that come to replace jobs but that they can use them to improve journalistic routines to limits never before reached. This union between journalists and machines (Túñez-López et al. 2018) revives the old debate that newsrooms should have new professional profiles and specialised work teams that connect the possibilities of artificial intelligence with the needs of journalism itself. In short, there are the so-called exo-journalists (Tejedor and Vila (2021)), who know the computer language and are endowed with heterogeneous technical and linguistic skills that allow them to document, verify and generate content from a transmedia logic and from different approaches.

This study has some limitations. The analysis of the reports that were ultimately posted could have more precisely quantified the percentage of improvement in each dimension studied, but that does not diminish the validity of the tool. Moreover, this study opens up new avenues for research comparing texts produced by humans and machines. An investigation into how the residents of depopulated areas view this type of information and its relevance to their lives would also be worth pursuing.

Author Contributions: Conceptualization, methodology, L.M.C.R.; formal analysis, investigation, writing, review and editing, L.M.C.R., M.J.U.R. and F.J.M.V. All authors have read and agreed to the published version of the manuscript.

Funding: This article was written under the auspices of the research project ‘Desigualdades informativas: cartografía de los desiertos mediáticos y hábitos de consumo en las zonas escasamente pobladas y en riesgo de despoblación de Castilla- La Mancha’ (SBPLY/23/180225/000051), financed by the European Regional Development Fund (ERDF). It also forms part of the research project ‘Inteligencia artificial y Periodismo: contenidos, audiencias, retos y desarrollo curricular (2023-GRIN-34286)’, financed by the University Research Plan and 85% co-financed by the European Regional Development Fund (ERDF).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Note

¹ These reports can be consulted on <https://www.rtve.es/elecciones-municipales-2023> (accessed on 20 March 2024).

References

- Aramburu Moncada, Luisa Graciela, Isaac López Redondo, and Antonio López Hidalgo. 2023. Inteligencia artificial en RTVE al servicio de la España vacía. Proyecto de cobertura informativa con redacción automatizada para las elecciones municipales de 2023. *Revista Latina de Comunicación Social* 81: 1–16. [CrossRef]
- Armentia Vizuete, José Ignacio, and José María Caminos Marcet. 2003. *Fundamentos de Periodismo Impreso*. Barcelona: Ariel Comunicación.
- Aydın, Ömer, and Enis Karaarslan. 2022. OpenAI ChatGPT generated literature review: Digital twin in healthcare. In *Emerging Computer Technologies*. Edited by Ö. Aydın. Izmir: Izmir Akademi Dernegi, pp. 22–31. [CrossRef]
- Brennen, J. Scott, Philipp N. Howard, and Rasmus K. Nielsen. 2022. What to expect when you’re expecting robots: Futures, expectations, and pseudo-artificial general intelligence in UK news. *Journalism* 23: 22–38. [CrossRef]
- Calvo-Rubio, Luis Mauricio, and José Luis Rojas-Torrijos. 2024. Criteria for journalistic quality in the use of artificial intelligence. *Communication & Society* 37: 247–59. [CrossRef]
- Calvo-Rubio, Luis Mauricio, and María José Ufarte-Ruiz. 2020. Percepción de docentes universitarios, estudiantes, responsables de innovación y periodistas sobre el uso de inteligencia artificial en periodismo. *El Profesional de la Información* 29: e290109. [CrossRef]
- Canavilhas, João. 2015. Nuevos medios, nuevo ecosistema. *El Profesional de la Información* 24: 357–62. [CrossRef]
- Carlson, Matt. 2015. The Robotic Reporter. Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism* 3: 416–31. [CrossRef]

- Clerwall, Christer. 2014. Enter the Robot Journalist. Users' perceptions of automated content. *Journalism Practice* 8: 519–31. [CrossRef]
- Cobo, Silvia. 2012. *Internet Para Periodistas. Kit de Supervivencia Para la era Digital*. Barcelona: UOC.
- Díaz-García, José Ángel, M. Dolores Ruiz, and María J. Martín-Bautista. 2020. NonQuery-Based Pattern Mining and Sentiment Analysis for Massive Microblogging Online Text. *IEEE Access* 8: 78166–82. [CrossRef]
- Digiday. 2017. Spanish Publisher El País Drove nearly 1000 Bot Subscribers over French Election. *Digiday*. Available online: <https://digiday.com/media/spanish-publisher-el-pais-drove-nearly-1000-news-bot-subscribers-french-election/> (accessed on 22 February 2024).
- Direito-Rebollal, Sabela, and Karen Donders. 2023. Public service media as drivers of innovation: A case study analysis of policies and strategies in Spain, Ireland, and Belgium. *Communications* 48: 43–67. [CrossRef]
- Dovifat, Emil. 1959. *Periodismo*. Ciudad de México: Unión Tipográfica Editorial Hispanoamericana.
- Estatuto de Información de la Corporación RTVE. 2008. Available online: https://www.rtve.es/contenidos/corporacion/Estatuto_de_informacion.pdf (accessed on 14 May 2008).
- Fanta, Alexander. 2017. *Putting Europe's Robots on the Map: Automated Journalism in News Agencies*. Oxford: University of Oxford, Reuters Institute for the Study of Journalism.
- Fieiras-Ceide, César, Martín Vaz-Álvarez, and José Miguel Túniz-López. 2023. Designing personalisation of European public service media (PSM): Trends on algorithms and artificial intelligence for content distribution. *Profesional de la Información* 32: 1–13. [CrossRef]
- Grijelmo, Álex. 2014. *El Estilo del Periodista*. London: Taurus.
- Guerrero-Solé, Frederic, and Coloma Ballester. 2023. El impacto de la Inteligencia Artificial Generativa en la disciplina de la comunicación. *Hipertext.net* 26: 1–3. [CrossRef]
- Guida, Giovanni, and Giancarlo Mauri. 1986. Evaluation of natural language processing systems: Issues and approaches. *Proceedings of the IEEE* 74: 1026–35. [CrossRef]
- Guía de Igualdad de RTVE. 2022. Available online: https://www.rtve.es/contenidos/documentos/guia_igualdad_2020.pdf (accessed on 8 March 2024).
- Haim, Mario, and Andreas Graefe. 2017. Automated News. Better than expected? *Digital Journalism* 5: 1044–59. [CrossRef]
- Hatcher-Moore, Philip. 2023. Salco 2023: Making News Hyperlocal. *BBC*. Available online: <https://www.bbc.co.uk/rnewslabs/news/salco-2023> (accessed on 19 May 2023).
- Hayes, Andrew F., and Klaus Krippendorff. 2007. Answering the Call for a Standard Reliability Measure for Coding Data. *Communication Methods and Measures* 1: 77–89. [CrossRef]
- Jia, Chenyan, and Thomas Johnson. 2021. Source Credibility Matters: Does Automated Journalism Inspire Selective Exposure? *International Journal of Communication* 15: 3760–81.
- Jung, Jaemin, Youngju Kim, Haeyeop Song, Hyunsuk Im, and Sewook Oh. 2017. Intrusion of software robots into journalism: The public's and journalists' perceptions of news written by algorithms and human journalists. *Computers in Human Behavior* 71: 291–98. [CrossRef] [PubMed]
- Krippendorff, Klaus. 2004. Measuring the Reliability of Qualitative Text Analysis Data. *Quality & Quantity* 38: 787–200. [CrossRef]
- LeCompte, Celeste. 2015. Automation in the Newsroom. *How algorithms can help reporters*. *Nieman Reports* 69: 32–45. Available online: <http://niemanreports.org/wp-content/uploads/2015/08/NRsummer2015.pdf> (accessed on 23 March 2024).
- Lermann Henestroza, Angelica, Hannah Grievling, and Joachim Kimmerle. 2023. Automated journalism: The effects of AI authorship and evaluative information on the perception of a science journalism article. *Computers in Human Behavior* 138: 107445. [CrossRef]
- Lopezosa, Carlos. 2023. ChatGPT y comunicación científica: Hacia un uso de la Inteligencia Artificial que sea tan útil como responsable. *Hipertext.net* 26: 17–21. [CrossRef]
- Lopezosa, Carlos, Carles Pont-Sorribes, Lluís Codina, and Mari Váñez. 2023. Use of generative artificial intelligence in the training of journalists: Challenges, uses and training proposal. *Profesional de la Información* 32: e320408. [CrossRef]
- López Delacruz, Santiago. 2023. Un vínculo paradójico: Narrativas audiovisuales generadas por inteligencia artificial, entre el pastiche y la cancelación del futuro. *Hipertext.net* 26: 31–35. [CrossRef]
- López Hidalgo, Antonio. 2001. *El Titular: Manual de Titulación Periodística*. Salamanca: Comunicación Social Ediciones y Publicaciones.
- López Hidalgo, Antonio. 2019. *El Titular. Tratado Sobre las Técnicas, Modalidades y Otros Artificios Propios de la Titulación Periodística*. Salamanca: Comunicación Social Ediciones y Publicaciones.
- Mandato-Marco a la Corporación RTVE previsto en el artículo 4 de la Ley 17/2006, de 5 de junio, de la Radio y la Televisión de Titularidad Estatal, aprobado por los Plenos del Congreso de los Diputados y del Senado. 2008. Available online: <https://www.boe.es/boe/dias/2008/06/30/pdfs/A28833-28843.pdf> (accessed on 12 December 2007).
- Manual de Estilo de RTVE. 2010. Available online: <https://manualdeestilo.rtve.es> (accessed on 23 July 2010).
- Martínez Albertos, J. L. 1974. *Redacción Periodística. (Los Estilos y los Géneros en la Prensa Escrita)*. Barcelona: A.T.E.
- Murcia Verdú, Francisco José, Rubén Ramos Antón, and Luis Mauricio Calvo Rubio. 2022. Análisis comparado de la calidad de crónicas deportivas elaboradas por inteligencia artificial y periodistas. *Revista Latina de Comunicación Social* 80: 91–111. [CrossRef]
- Odrizola-Chéné, Javier, Javier Díaz Noci, Ana Serrano-Tellería, Rosa Pérez-Arozamena, Laura Pérez-Altable, Juan Linares-Lanzman, Lucía García-Carretero, Luis Mauricio Calvo-Rubio, Manuel Torres-Mendoza, and Adolfo Antón-Bravo. 2020. Inequality in times of pandemics: How online media are starting to treat the economic consequences of the coronavirus crisis. *Profesional de la Información* 29: e290403. [CrossRef]

- OpenAI. 2022. Introducing ChatGPT. Available online: <https://openai.com/blog/chatgpt> (accessed on 23 July 2010).
- Rojas Torrijos, José Luis. 2021. Semi-Automated Journalism: Reinforcing Ethics to Make the Most of Artificial Intelligence for Writing News. In *News Media Innovation Reconsidered*. Edited by María Luengo and Susana Herrera. Hoboken: Wiley-Blackwell, pp. 124–37.
- Rojas Torrijos, José Luis, and Carlos Toural-Bran. 2019. Periodismo deportivo automatizado. Estudio de caso de AnaFut, el bot desarrollado por El Confidencial para la escritura de crónicas de fútbol. *Doxa Comunicación* 29: 235–54. [CrossRef]
- RTVE. 2023. RTVE Pone a Prueba la Inteligencia Artificial Para Cubrir las Elecciones del 28 de mayo. *RTVE.es*. Available online: <https://www.rtve.es/noticias/20230528/inteligencia-artificial-noticias-elecciones-28-mayo-poblaciones-menos-1000-habitantes/2446742.shtml> (accessed on 23 July 2010).
- Salaverría, R. 2005. *Redacción Periodística en Internet*. Pamplona: EUNSA.
- Sánchez Gonzales, Hada M., and María Sánchez González. 2017. Los bots como servicio de noticias y de conectividad emocional con las audiencias. El caso de Politibot. *Doxa Comunicación* 25: 63–84. [CrossRef]
- Sandoval-Martín, Teresa, and Leonardo La-Rosa Barrolleta. 2023. Investigación sobre la calidad de las noticias automatizadas en la producción científica internacional: Metodologías y resultados. *Cuadernos.info* 55: 114–36. [CrossRef]
- Tandoc, Edson C., Jr., Lim Jia Yao, and Shangyuan Wu. 2020. Man vs. Machine? The Impact of Algorithm Authorship on News Credibility. *Digital Journalism* 8: 548–62. [CrossRef]
- Tejedor, Santiago. 2023. *La Inteligencia Artificial en el Periodismo: Mapping de Conceptos, Casos y Recomendaciones*. Barcelona: Editorial UOC.
- Tejedor, Santiago, and Pere Vila. 2021. Exo Journalism: A Conceptual Approach to a Hybrid Formula between Journalism and Artificial Intelligence. *Journalism and Media* 2: 48. [CrossRef]
- Trillo-Domínguez, Magdalena, and Jordi Alberich-Pascual. 2017. Deconstrucción de los géneros periodísticos y nuevos medios: De la pirámide invertida al cubo de Rubik. *El Profesional de la Información* 26: 1091–99. [CrossRef]
- Túñez-López, José Miguel, Carlos Toural-Bran, and Santiago Cacheiro-Requeijo. 2018. Uso de bots y algoritmos para automatizar la redacción de noticias: Percepción y actitudes de los periodistas en España. *El Profesional de la Información* 27: 750–58.
- Ufarte Ruiz, María José, and Francisco José Murcia Verdú. 2018. Desarrollo académico y profesional sobre el uso de la inteligencia artificial en las redacciones periodísticas. *Textual & Visual Media* 11.
- Van-Dis, Eva A. M., Willem Zuidema, Johan Bollen, van Robert Rooij, and Claudi L. Bockting Alo. 2023. ChatGPT: Five priorities for research. *Nature* 614: 224–26. [CrossRef] [PubMed]
- Waddell, T. Franklin. 2019. Attribution Practices for the Man-Machine Marriage: How Perceived Human Intervention, Automation Metaphors, and Byline Location Affect the Perceived Bias and Credibility of Purportedly Automated Content. *Journalism Practice* 13: 1255–72. [CrossRef]
- Wang, Shuai, Bevan Koopman, Harrisen Scells, and Guido Zuccon. 2023. Can ChatGPT Write a Good Boolean Query for Systematic Review Literature Search? *arXiv* arXiv:2302.03495. doi:10.48550/arXiv.2302.03495. [CrossRef]
- Wölker, Anja, and Thomas E. Powell. 2018. Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism* 22: 1–18. [CrossRef]
- Zaragoza, Fuster María Teresa, and José Alberto García Avilés. 2022. Public Service Media laboratories as communities of practice: Implementing innovation at BBC News Labs and RTVE Lab. *Journalism Practice* 18: 1256–74. [CrossRef]
- Zheng, Yue, Bu Zhong, and Fan Yang. 2018. When algorithms meet journalism: The user perception to automated news in a cross-cultural context. *Computers in Human Behavior* 86: 266–75. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.