GENERATIVE ARTIFICIAL INTELLIGENCE AND MEDIA TRUST. AN ANALYSIS OF AI DETECTION IN NEWS USING GPTZERO

🆔 **López del Castillo Wilderbeek, Francisco Leslie[1]:** Pompeu Fabra University. Spain.
franciscoleslie@alumni.upf.edu

How to cite the article:

ABSTRACT

Introduction: The use of AI in news generation is a reality that can jeopardize trust in the media. **Methodology:** The aim of this article was to determine through content analysis whether the best platform for detecting the use of AI in texts: GPTZero, is guaranteed to distinguish between news that are completely created with AI and contents signed by professionals. For this purpose, texts of real news (press and digital) and news developed by artificial intelligence (ChatGPT) were analyzed. **Results:** The results obtained indicated that GPTZero offers a low reliability for detecting texts generated with ChatGPT (7.3%) and that it can signal as false positives contents that have not been developed with artificial intelligence. **Discussion:** The analyses performed with two different algorithms indicated that GPTZero did not discover the full use of ChatGPT in a corpus of news and, likewise, warned of a certain probability of collaborative use of AI (writing by a human with AI support) in news that were signed by professional journalists. **Conclusions:** The results obtained have generated more doubts than certainties about whether GPTZero can help signal the use of AI or ensure the prestige of the media.

Keywords: media, artificial intelligence, GPTZero, ChatGPT, press, cyber media.

---

[1] **Francisco Leslie López del Castillo Wilderbeek**: PhD in Communication from the Universitat Pompeu Fabra, Collaborating Professor at the UOC, Associate Professor at the UOC. He works as a media analyst at Rebold.

1. INTRODUCTION

1.1. Artificial Intelligence and News Reporting

As warned by the Reuters Institute, media outlets are in the process of integrating Artificial Intelligence into their products. Specifically, 28% of the media outlets in the survey state that they are already working with this type of technology, while 39% are testing its inclusion (Newman, 2023).

A very relevant variant within the framework of AI is generative AI, since this technology can create content such as text, images or even programming code based on prior knowledge and thanks to a few brief instructions. In this way, the generation of plausibly novel content represents a very significant leap, since it can be confused with human-created material, which is indeed original material.

Applying this technology in news reporting has become a fact for media outlets as it leads to cost savings (Aramburú Moncada et al., 2023) despite the need to hire specialized personnel (de-Lima-Santos and Salaverría, 2021). However, it also fosters an added problem since news is a public good in a way that may alert regulatory agencies (Sandrini & Somogyi, 2023). Similarly, and disregarding the supposed neutrality inherent in the role of the media, the refinement of these processes can lead to the proliferation of deepfakes (highly credible fake images) and, most especially, the dissemination of misinformation (Shoaib et al., 2023). This is because generative AI, trained with large textual models, can create fake news without any difficulty, which is coherent (Botha & Pieterse, 2020) and with a style almost identical to that of authentic professionals even though the result is not necessarily truthful (Franganillo, 2023). According to a survey conducted by IPSOS in 29 countries 74% of the population believes that AI makes it easier to generate credible fake news (Dunne, 2023).

In addition to this, the work of Longoni et al. (2022) concluded that the acknowledged use of AI by news producers can be detrimental even if the content is true. According to these authors if media outlets disclose that they use AI for their content they are damaging the trust and credibility that recipients have in news producers, even if the news is truthful. Specifically, they went so far as to state that "an important consequence of our experiments is that calls for transparency in the use of AI can be counterproductive" (Longoni et al., 2022, p. 102).

Therefore, the use of AI by news producers generates a double danger. On the one hand, generative AI is strongly related to the mass production of fake news. On the other hand, audience acceptance is not the same if they are aware that they are facing products that have been produced, in whole or in part, with the help of AI.

1.2. Detecting AI-generated news

The use of AI for news production is controversial. To mitigate the impact of AI-generated fake news, there are several proposals that are in line with the processes already used in detecting manually generated fake news.

According to Shoaib et al. (2023), the need to develop more sophisticated technological solutions, to develop international policies that are flexible to the technological environment and to promote media literacy are of vital importance. Regarding the volume of fake news production by AI, Deepak (2023) states that there is a need for an improvement of automation in fake news detection at the same level as generative AI improves its production processes. The use of applications for AI detection would transcend or complement the human verification function (fact-checking) and, in some cases, avoid subjective biases (Rodríguez-Pérez et al., 2023) where this technology has been used to mislead audiences. Although it should be noted that this would not be fact-checking in the strict sense, since this process is related to the "ability to detect claims, retrieve relevant evidence, evaluate the veracity of each claim, and provide justification for the conclusions provided" (DeVerna et al., 2023, p. 2). What AI technology can offer is to reveal whether it has been used, either legitimately or illegitimately, but not to replace human cognition, judgment and emotional intelligence.

However, beyond the detection of the use of AI in news that have actually used this technology as a support, there is the underlying question of whether a content developed with AI by a reputable producer (for example, a media outlet that also publishes offline) is negatively affected by the audience's perception of the use of AI in the news it consumes.

For this reason, applications or programs dedicated to uncovering the use of AI for news reporting could damage the reputation of media outlets that use this technology. Moreover, in cases where the use of AI is incorrectly detected (false positive) the affected media could be falsely accused and, therefore, presumably lose the trust of their audience as stated in the work of Longoni et al. (2022).

It is no coincidence that according to the results of the I Study on disinformation in Spain (Sádaba-Chalezquer & Salaverría-Aliaga, 2022) 84.6% of respondents showed a preference for traditional media to get information to the detriment of social networks because the former have "professional teams of journalists who verify, contrast and analyze information" (Sádaba-Chalezquer & Salaverría-Aliaga, 2022, p. 11). However, the issue would be that this trust would be strongly damaged if the participation of qualified professionals is minimized or directly disappeared by the advent of generative AI.

Currently, the reference among the applications available for detecting the use of AI in textual corpora is the GPTZero program. As stated on its website, this program is "the gold standard in AI detection, trained to detect ChatGPT, GPT4, Bard, LLaMa and other AI models" (GPTZero, s.f.).

2. OBJECTIVES

The aim of this research is to elucidate whether GPTZero is really a guarantee to detect AI-generated fake news (news completely elaborated without being based on a real event.) It also aims to determine the use of AI as a support for news producers and/or

whether it poses a threat to traditional media by detecting signs of AI use that may jeopardize their credibility as a reference media.

Specifically, the aim is to analyze the hit rate of GPTZero on textual corpora. On the one hand, there are actual news extracted from reference media (both print and digital press) and on the other hand, contents that are generated by ChatGPT artificial intelligence.

3. METODOLOGÍA

To fulfill the objective stated above, this research has analyzed a corpus of pieces of news using GPTZero. This corpus has been divided into two groups. On the one hand, news items entirely generated by the AI Chat-PGT 3.5 by means of the instruction "Elaborate a news item about (NAME OF ORGANIZATION)". The choice of this instruction was because ChatGPT does not reply to the instruction "Elaborate a fake news about (NAME OF ORGANIZATION)". In the same way, this would be an invented news (not based on real events) generated by AI (fake news). To confirm this, this corpus of news items was analyzed manually, and it was confirmed that the facts narrated by ChatGPT are not linked to real present or past events. Specifically, the organizations for which fake news was generated by ChatGPT were Iberdrola, Naturgy and Acciona.

On the other hand, a corpus of news items published in both the print and digital press about the three organizations mentioned above was collected (one news item for each organization and each media outlet). The media that were analyzed in the traditional press were *Expansión*, *Cinco Días* and *El Economista*, as they are the economic media of reference in Spain. In the case of digital media, only native digital media (without offline matrix) were chosen: *El Confidencial, OK Diario, El Español* and *Vozpópuli*. All of them are among the five best native digital media according to the ranking of the media portal PR Noticias, dated October 2023 (PR Noticias, n.d.).

In both corpora, the name of the three organizations already mentioned in the headline (to ensure the prominence of the organization within the analyzed news item) was used as a search expression to identify them.

The contents of the three corpora (IA, paper and digital) were run through the GPTZero analyzer on January 12, 2023. This is an important fact since the AI detection algorithm was updated on January 17, 2024, so that two different results were obtained by replicating the same experiment and the same conditions.

From these three corpora, in addition to the analysis by GPTZero, variables in the 24 units of analysis were selected by content analysis to look for correlations with the percentage of possibility of AI intervention. The analysis categories obtained were: number of words, presence of the organization entity, journalist's signature (if applicable) and type of source (digital native or traditional paper).

As Bardin (1991) points out, content analysis makes it possible to obtain indicators, in this case both quantitative and qualitative ones. On the one hand, the quantitative

indicators were the volume of total words and the volume of the key word of the entity. On the other hand, the qualitative indicators were the journalist's signature (dichotomous variable in which a condition is given or not given) and type of source (nominal qualitative variable).

Table 1

Table with the applied methodology for the content analysis

Analysis category	Indicator	Type of variable
Text volume	Quantitative	Quantitative
Keyword volume for the entity	Quantitative	Quantitative
Journalist's signature	Qualitative	Dichotomous variable
Type of source	Qualitative	Nominal qualitative variable

Source: Elaborated by the authors.

4. RESULTS

4.1. First GPTZero analysis (January 12, 2024)

As mentioned above, the three content corpora were analyzed using GPTZero. The results about AI involvement in generating news items were expressed in the form of percentage of possibilities that the text had been elaborated exclusively by applying generative AI.

Table 2

GPTZero analysis results January 12, 2024

Entity	Chat GPT	Cinco Días	Expansión	El Economista	El Confidencial	OK Diario	Vozpópuli	El Español
Iberdrola	9%	5%	9%	5%	5%	5%	5%	9%
Naturgy	5%	5%	5%	5%	5%	5%	5%	5%
Acciona	8%	5%	7%	5%	7%	6%	7%	6%

Note: *Probability according to GPTZero (algorithm before January 17, 2024) indicating that the analyzed content was entirely produced by AI. Content was segmented according to the entity featuring the news item.*

Source: *Elaborated by the authors with results obtained by GPTZero (before January 17, 2024 algorithm).*

As it can be seen in Table 2, the results obtained with the pre-January 17 algorithm are not very different from each other. Specifically in the case of news items about Naturgy (both fake and real) GPTZero considered that all of them had the same possibilities of having been developed by AI. In the other examples, the percentages ranged from the same situation found in Naturgy (same percentage of original content with respect to ChatGPT content) to lower values for AI-created material (higher level of human participation).

López del Castillo Wilderbeek, Francisco Leslie.

Generative artificial intelligence and media trust. An analysis of AI detection in news using GPTZero

It is particularly significant that, while the AI corpus was generated exclusively with ChatGPT by means of a simple instruction, it was not adequately detected by GPTZero, which showed a maximum result of 9% chance that the material was generated exclusively with AI when this was precisely the case. On the other hand, the minimum percentage was 5%, which is even further from reality (even though ChatGPT 3.5 is one of the technologies recognized by GPTZero, as the platform specifically states).

On the other hand, in the case of false positives (detecting full use of AI in texts that have not been recognized as such by the publisher) the analysis prepared by GPTZero was not conclusive in the sense that in one of the three typologies of news (Iberdrola, Naturgy and Acciona) the same levels of reliability were found about AI usage in comparison with the content that really had been generated entirely by AI. This comparison between truthful content and artificially produced content further raises, once again, concerns about the reliability of the application.

No relevant changes were found between print and native digital media, so there was no evidence of any kind of prevalence of one channel over another when it came to detecting the presence or absence of AI technology.

More specifically, the average chance of exclusive AI generation was 5.6% in the case of print media and 5.83% in native digital media. There was no significant change in the ratio from one media to another, although two print media showed the lowest results for the possibility of exclusive AI generation (5% in both cases).

4.2. Second GPTZero analysis (January 12, 2024)

The changes made in GPTZero on January 17, 2024, resulted in an alteration of the results obtained in the analysis of the texts. More specifically, the platform provided a more detailed result that not only expressed the probability that the content had been generated entirely by AI, but also the probability that it was content developed jointly by a human and an artificial intelligence. In this way, the algorithm provided information on three variables: fully human content creation (human), a combination of human and AI (mixed) and finally, fully AI content creation (ai), all expressed in percentages as in the previous algorithm.

Taking these changes into account, the three textual corpora were reanalyzed.

Table 3

GPTZero Analysis Results January 20, 2024

Entity	ChatG PT		Cinco Días		Expans ión		El Econo mista		El Confid encial		OK Diario		Vozpó puli		El Español	
Iberdrola	88	12	90	10	94	6	92	8	81	19	89	11	81	19	88	12
Naturgy	84	16	90	10	87	13	89	11	88	12	90	10	90	10	89	11
Acciona	84	16	89	11	92	8	87	13	87	13	88	12	87	13	81	18

Note*: Probability according to GPTZero (algorithm post January 17, 2024) indicating that the analyzed content was entirely developed by a human being or by AI. The results concerning fully AI developed content were in all cases equal to 0%. Content was segmented according to the entity featuring the news item.*
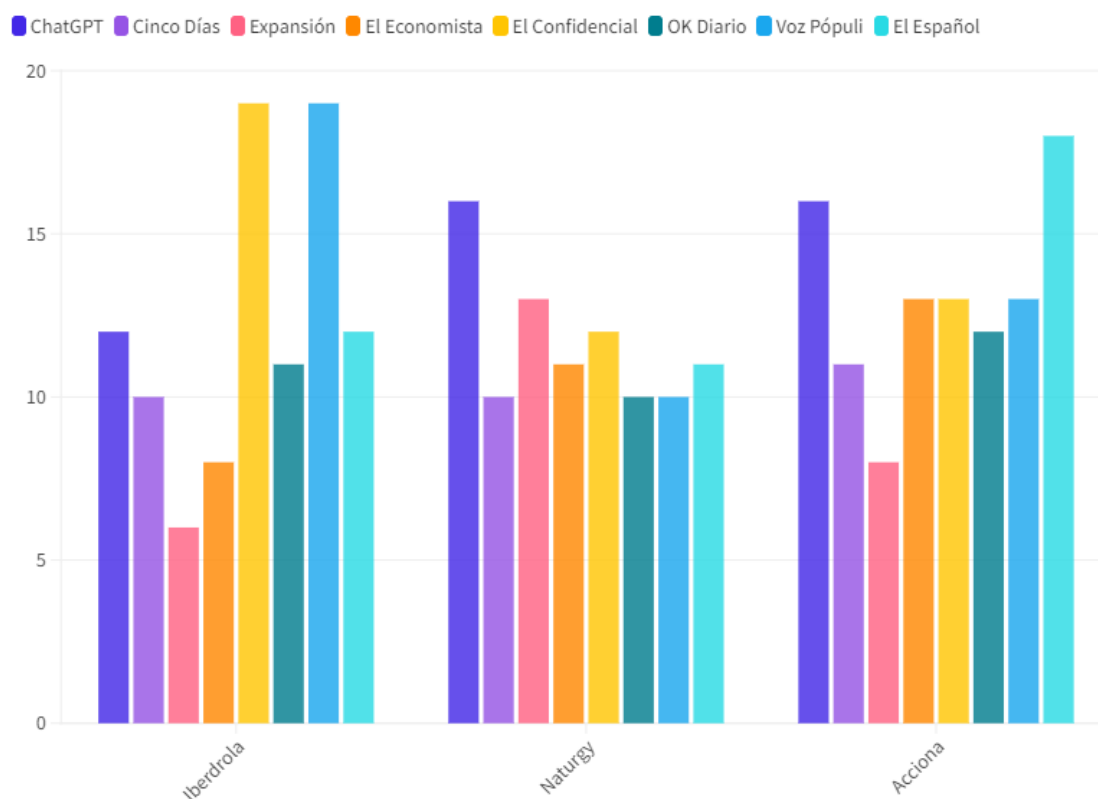
Source*: Elaborated by the authors with results obtained by GPTZero (after January 17, 2024, algorithm).*

As it can be seen in Table 3, more detail was provided in the results obtained. An important element is that none of the contents was found to have exclusive AI intervention (contrary to the previous algorithm) and therefore all the texts were considered reliable even though three of them were automatically generated with ChatGPT.

In turn, compared to the previous algorithm the hit rate was even lower when considering how likely it was that the content was collaboratively developed by human/AI. This is because GPTZero (post algorithm) found that the material being developed with ChatGPT had a 12% and 16% (in two cases) chance of being a collaboratively developed material. This, as mentioned above, is a false negative as it was not collaborative material but theoretically original human material. However, it may be an indicator as high levels of the mixed variable can be considered signals of AI generation. However, the results on human/AI collaboration were lower in the ChatGPT texts than in the stand-alone texts as it can be seen in Figure 1.

Figure 1

Probability of AI-generated content based on GPTZero analysis results, January 20, 2024.

Source: Elaborated by the authors with data obtained from GPTZero.

Three cases were found in which GPTZero detected only an 81% chance that the text was fully human when in ChatGPT content the lowest probability was 84%. In other words, while the previous algorithm found an erroneous match (similar levels) this new algorithm went so far as to give more margin of plausibility to AI-created material than to texts that were presumably the work of communication professionals.

In the end, with respect to the average data, ChatGPT content offered an average of 85.3% chance that it was material created by human-machine collaboration. In print news the average increased to 90% and in digital news from native media the average was 86.58%. As can be seen, the differences were minimal, especially when comparing digital news with news generated by ChatGPT (86.58% vs. 85.3%).

Analyzing the data for each media outlet, there was no evidence that a particular media outlet dragged down or boosted the overall results with respect to AI participation in the newsroom. In the print press, the differences by newspaper were insignificant: *Cinco Días* reached 89.6%, *Expansión* 90% and *El Economista* 89.3%.

Within the native digital media, although it is true that *OK Diario* achieved a low AI participation ratio and close to the results of the print press (89% on average for an 11% average probability of being fully developed by AI), the rest of the digital media were in similar ranges: *El Confidencial* 85.3%, *Vozpópuli* 86% and *El Español* 86%. One

relevant detail is that the three news items in *OK Diario* were signed by professional journalists (Carlos Ribagorda, Benjamín Santamaría and Eduardo Segovia) while a few of the digital news corpus had their origin in news agencies. In *El Español* the news item about Iberdrola was signed with the expression Agencias and in the case of El Confidencial, one was signed by EUROPA PRESS (Naturgy company and 12% probability of being the result of AI) and another by EFE (Acciona company and 13% probability of being the result of AI). The most characteristic element would be, however, that in the subcorpus of articles retrieved from *El Confidencial*, the articles with the lowest probability ratios of being generated entirely by AI were those in which the authorship was by an agency, as opposed to the article with the highest probability of being generated by AI (19%) which, contrary to general intuition, was signed by a journalist.

4.3. Content analysis of news items less likely to be entirely generated by humans

To understand in greater detail, the results obtained in point 3.2, content analysis was applied to the three news items that were identified as being more likely to have been generated by an AI in collaboration with a human (even more than the three sample items that actually met this requirement).

Content analysis allows the understanding of a text in an objective and systematic way by quantifying the elements that are characteristic of it according to predetermined criteria (Neuendorf, 2017).

The three news items were:

1) "Iberdrola renuncia a comprar PNM, la gran operación que EEUU vetó por el caso Villarejo" [Iberdrola refuses to buy PNM, the large operation that the US vetoed due to the Villarejo case]

El Confidencial / Iberdrola / January 2, 2024

Possibility of being generated by AI/human collaboration: 81%.

Possibility of being generated entirely by AI: 19%

https://www.elconfidencial.com/empresas/2024-01-02/iberdrola-fusionar-avangrid-pnm-resources_3803958/

2) "Iberdrola eleva un-12,2% el dividendo a cuenta y amplía capital en 1.304 millones" [Iberdrola increases its interim dividend by 12.2% and capital is raised by 1,304 million]

Vozpópuli / Iberdrola / January 5, 2024

Possibility of being generated by AI/human collaboration: 81%.

Possibility of being generated entirely by AI: 19%

López del Castillo Wilderbeek, Francisco Leslie.

Generative artificial intelligence and media trust. An analysis of AI detection in news using GPTZero

https://www.vozpopuli.com/economia_y_finanzas/empresas/iberdrola-eleva-dividendo-cuenta-amplia-capital-1304-millones.html

3) "Acciona anuncia que construirá una planta pionera de reciclado de palas eólicas en Navarra operativa en 2025" [Acciona announces that it will build a pioneering wind blade recycling plant in Navarra that will be operational in 2025]

El Español / Acciona / November 15, 2023

Possibility of being generated by AI/human collaboration: 81%.

Possibility of being generated entirely by AI: 18%[2]

https://www.elespanol.com/invertia/empresas/energia/20231115/acciona-anuncia-construira-planta-pionera-reciclado-palas-eolicas-navarra-operativa/809919384_0.html

The comparison between these three news items does not allow, in the first instance, to draw any correlations about their relevance to the group of content most likely to have been generated by AI in collaboration with a human being.

On the one hand, there is no uniformity regarding the protagonist entities (although there are two news items with Iberdrola as the main entity). On the other hand, there is a variety of media and length (*El Confidencial* 814 words, *Vozpópuli* 474 and *El Español* 773) and presence of the key word entity (the term Iberdrola appeared 8 times in *El Confidencial*, 4 times in *Vozpópuli* and the term Acciona 6 times in *El Español*). In turn, there was a certain balance between the presence of the keyword entity and the length of each text (in *El Confidencial* Iberdrola appeared with an average of 101.75 words, in Vozpópuli with an average of 118.5 words and Acciona with an average of 128.8 words).

The most significant conditioning factor could be that they are all digital content. Considering that 12 were analyzed, it can be affirmed that 25% of these news items were detected as potentially generated by AI (in comparison with the total corpus and taking into account that in no case did GPTZero venture to point out contents as exclusive to AI, but rather with low levels of "humanity").

However, the content analysis also evaluated the authorship of the information, detecting that all of them were signed by professionals of the sector. *El Confidencial* was signed by Miquel Roig and Javier Melguizo, *Vozpópuli* by Fernando Asunción and *El Español* by Laura Ojea.

This data is relevant in the sense that these reports, presumably with a higher rate of participation of AI, enjoyed the support and prestige not only of a media outlet but of

---

[2] In this particular case GPTZero identified an 81% chance of being human-made material, an 18% chance of being made in AI/human collaboration, and yet the remaining 1% that would complete the 100% did not appear (the chance of full AI use was 0%).

a group of professionals (since there is the possibility of signing the texts with the generic term such as newsroom or agency).

4.4. Content analysis of news items most likely to be completely human generated

To understand in depth, the results obtained in point 3.2, content analysis was applied to the three news items that were identified as most likely to have been generated exclusively by a human being.

1) "Iberdrola se lanza con BP a por 12.000 electrolineras con mil millones" [Iberdrola and BP are working together to invest one billion euros in 12,000 electric gas stations]

Expansión Page 3 / December 2, 2023

Possibility of being generated by AI/human collaboration: 94%

Possibility of being generated entirely by AI: 6%

2) "Iberdrola lanza su planta de amoniaco en Portugal" [Iberdrola launches its ammonia plant in Portugal]

El Economista Page 6 / January 12, 2024

Possibility of being generated by AI/human collaboration: 92%

Possibility of being generated entirely by AI: 8%

3) "Acciona nombra CEO de Infraestructuras a José Díaz-Caneja" [Acciona appoints José Díaz-Caneja as CEO of Infrastructure]

Expansión Page 10 / December 15, 2023

Possibility of being generated by AI/human collaboration: 92%

Possibility of being generated entirely by AI: 8%

The comparison between these three news items does not allow, once again, to draw correlations about their relevance to the group of contents with greater possibilities of having been generated by humans without the support of AI technology. On the one hand, the length of the texts was heterogeneous (355 for Iberdrola in *El Economista*, 327 for Iberdrola in *Expansión* and 394 for Acciona in *Expansión*). On the other hand, these lengths showed no significant discrepancy with the average of the entire corpus of the print press, which was 346.5 words per news item.

Similarly, the presence of the term protagonist entity (Iberdrola and Acciona) within the text was also uneven, with only 3 references to Iberdrola in *El Economista*, 5 to Iberdrola in *Expansión* and 8 to Acciona in *Expansión*.

Thus, the analysis of the content failed to find an identifiable key element that could predetermine the detection of greater human participation. Likewise, the analysis by

GPTZero did show a very evident variable and that is that three native digital contents, even if signed, had higher rates of AI participation compared to three paper press contents that had the highest rates of human participation.

4.5. Correlation of number of words

As a last variable to be considered, a word count of the contents was carried out and correlations were identified between this data and the result offered by GPTZero segmented by channel (print or digital press). To determine the existence or not of a link between both variables, the Pearson correlation coefficient was applied, which is a statistical measure that allows understanding the relationship between two sets of data. The aim was to determine whether the volume of texts could determine whether GPTZero could detect the presence or absence of AI intervention in their production to a greater or lesser extent.

Table 4

Word count of content under analysis by GPZero

Entity	ChatG PT	Cinco Días	Expa nsión	El Econo mista	El Confid encial	OK Diario	Vozpó puli	El Español
Iberdrola	355	327	343	355	814	774	474	611
Naturgy	349	372	376	349	656	594	605	552
Acciona	364	394	360	243	303	557	773	498

Source: Elaborated by the authors

The analysis differentiated between news items that were published in print media and those published in native digital media. In digital media, no correlation was found (text volume does not affect the percentage of detecting AI intervention).

On the other hand, in print media, a weak correlation was found between the possibility that the text was completely generated by humans and the number of words (coefficient value 0.3087 when the value for perfect correlation would be 1). This means that the more words the text contains, the more chances there are that GPTZero will detect (correctly) that it is content generated by humans (in this case professional journalists). However, as indicated, the value obtained (0.3087) is not high enough to guarantee this, since in statistics an appropriate reliability occurs at values of 0.5 or higher.

However, this finding, although insufficient, could partially support the data previously observed, according to which press contents on paper are more likely to be recognized as material created by humans. In any case, it could be affirmed that there is a bias towards the detection of human participation in paper news, especially in those cases in which the writing is proportionally longer than its peers.

To confirm this, a student's t-test was performed to observe the statistical value of the difference in means between two groups. The groups that were analyzed were the number of words in each text according to the source (ChatGPT, *Cinco Días, Expansión,*

El Economista, El Confidencial, OK Diario, Vozpópuli and *El Español*) and the result obtained in GPTZero in each of these texts.

Results were as follows: ChatGPT 74.8 / Cinco Días 17.9 / *Expansión* 35.9 / *El Economista* 8.3 / *El Confidencial* 3.8 / *OK Diario* 9.4 / *Vozpópuli* 6.9 and *El Español* 16.5.

Although the least prominent case is the corpus of El Confidencial, it can be concluded that there is a significant difference between the means of the two groups (number of words/probabilities according to GPTZero). This is because when comparing the average values of two data sets, these averages are sufficiently different to conclude that this difference is not a mere coincidence. This data indirectly confirms with that obtained in the Pearson correlation coefficient according to which it cannot be affirmed that there is a cause/effect link between the number of words or the typology of the texts in relation to the possibilities of these being detected as presumably IA-generated material. However, in the same way, the author does not argue that print media have characteristics that differentiate their chances of being detected or not as AI-created material.

5. CONCLUSIONS

This article aimed to address a first-order question since the entry of AI in news production: Are there models that can efficiently detect the use of this technology in the material that the media produce?

The justification lies in the negative impact and mistrust that AI can generate in media audiences. On the one hand, generative AI is a very effective tool for the mass generation of fake news. On the other hand, recipient audiences have shown their doubts about the use of AI by news producers when they have been transparent about its use, even if it was as a support for human work (Longoni et al., 2022).

To meet this objective, an experiment was carried out by means of what is currently the canonical platform for generative AI detection in texts: GPTZero. Three corpora of texts (generated by ChatGPT, published in print and published in native digital media) were prepared and analyzed with the platform (one analysis with an algorithm before 17 January 2024 and one with the algorithm after that date).

The results were striking in the sense that the detection of AI in the texts produced by ChatGPT was very low using the first algorithm and even worse using the second algorithm. In other words, in the first analysis, the probability that the text was recognized as AI-generated was low (mean of 7.3%), especially considering that this data was not very far from the parameters obtained in media texts (mean of 5.76%).

The results for the second algorithm were, if possible, even worse, since the probability that the texts generated by ChatGPT were indeed the result of an AI averaged 85.3%. A high result that becomes an alarm signal once contextualized, since in native digital media signed by journalists a minimum level of 81% was reached in three cases. Therefore, GPTZero in its most updated version is not a guarantee to distinguish

between AI-generated material and presumably human-generated material (flagging false positives).

This issue (the presumed ineffectiveness of GPTZero) leads, however, to other findings of interest. The foremost of these findings is that with the new algorithm three print news stories provided the highest results for human involvement while three other native digital news stories showed (as discussed above) the lowest results.

Such a result may pose an added danger for news producers in digital media because, although no distinguishing features were found, it could represent a bias whereby the channel typology somehow predefines the probability that their publications are more likely to be flagged as AI-mediated material. However, this should be confirmed in research with larger corpora.

Another relevant finding, despite of its relative significance, was that statistically analyzing the correlation between the number of words and GPTZero results (Pearson's coefficient), there was a certain link between the length of the wording of the paper news and the probability of obtaining a high value of human participation according to the algorithm. The specific result provided a weak correlation (value 0.3087 when a strong correlation occurs with a value 0.5 or higher) according to which the more words the paper news item has, the higher value of human involvement it gets according to ChatGPT. Although the t-test to calculate the difference in means did not detect significant differences, this data leads us to reflect on whether the inherent weight of a traditional medium such as the press published in physical format goes beyond the mere halo of prestige and is a competitive advantage over digital products, even if they are signed, in the same way, by professionals in the sector.

Finally, it is essential to understand that the first premise of this research could not be confirmed. The current model for detecting AI participation in texts (GPTZero) does not guarantee the mission for which it has been designed since it could not identify as such three news items created by ChatGPT with a simple instruction. Moreover, this algorithm showed false positives in human/IA collaboration (mixed variable) in all the news items of the textual corpora. However, the implementation of AI in the media is not yet sufficiently strong to make it reasonable that all of them had AI involvement (especially when professional journalists signed the articles).

The results of this research have not only failed to provide reassurance to the public and the media but also raise new doubts as to whether the massive use of AI detection models can cause, in the same proportion, unjustified concerns among the receiving public that end up damaging the prestige and discipline of the media.

6. REFERENCES

Aramburú Moncada, L. G., López-Redondo, I., & López Hidalgo, A. (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. https://doi.org/10.4185/RLCS-2023-1550

Bardin, L. (1991). *Análisis de contenido* (Vol. 89). Ediciones Akal.

Botha, J., & Pieterse, H. (2020). Fake news and deepfakes: A dangerous threat for 21st century information security. In B. K. Payne, & H. Wu (Eds.), *ICCWS 2020 15th International Conference on Cyber Warfare and Security* (pp. 57-66). Academic Conferences and Publishing International Limited. http://tinyurl.com/3zzbpnte

Deepak, P. (2023). AI and Fake News: Unpacking the Relationships within the Media Ecosystem. *Communication & Journalism Research, 12*(1), 15-32. http://tinyurl.com/mr3k3d5a

DeVerna, M. R., Yan, H. Y., Yang, K.-C., & Menczer, F. (2023). Fact-checking information generated by a large language model can decrease news discernment. *arXiv*. https://doi.org/10.48550/arXiv.2308.10800

Dunne, M. (2023). Data Dive: Fake news in the age of AI. *IPSOS*. http://tinyurl.com/nbwfzvu9

Franganillo, J. (2023). La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos. *Methaodos. revista de ciencias sociales, 11*(2), 1-17. https://doi.org/10.17502/mrcs.v11i2.710

GPTZero (n.d.) *More than an AI detector. Preserve what's human*. Retrieved [January 14, 2025] from http://tinyurl.com/y7jsnz6s

de-Lima-Santos, M. F., & Salaverría, R. (2021). From data journalism to artificial intelligence: challenges faced by La Nación in implementing computer vision in news reporting. *Palabra Clave*, 24(3). https://doi.org/10.5294/pacla.2021.24.3.7

Longoni, C., Fradkin, A., Cian, L., & Pennycook, G. (2022, June). News from generative artificial intelligence is believed less. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)* (pp. 97-106). Association for Computing Machinery. https://doi.org/10.1145/3531146.3533077

Neuendorf, K. A. (2017). *The content analysis guidebook*. SAGE Publications, Inc. https://doi.org/10.4135/9781071802878

Newman, N. (2023). Journalism, media, and technology trends and predictions 2023. Reuters Insitute for the Study of Journalism. https://doi.org/10.5287/bodleian:NokooZeEP

PR Noticias (n.d.) *Ranking: Prensa*. Retrieved [October 24, 2023] from http://tinyurl.com/57u3vzaw

Rodríguez-Pérez, C., Seibt, T., Magallón-Rosa, R., Paniagua-Rojano, F. J., & Chacón-Peinado, S. (2023). Purposes, principles, and difficulties of fact-checking in Ibero-America: Journalists' perceptions. *Journalism Practice*, *17*(10), 2159-2177. https://doi.org/10.1080/17512786.2022.2124434

Sádaba-Chalezquer, C., & Salaverría-Aliaga, R. (2022). *I Estudio sobre la desinformación en España. Un proyecto de UTECA y la Universidad de Navarra*. http://tinyurl.com/9t5cxb62

Sandrini, L., & Somogyi, R. (2023). Generative AI and deceptive news consumption. *Economics Letters*, 232, 111317. https://doi.org/10.1016/j.econlet.2023.111317

Shoaib, M. R., Wang, Z., Ahvanooey, M. T., & Zhao, J. (2023). Deepfakes, Misinformation, and Disinformation in the Era of Frontier AI, Generative AI, and Large AI Models. In *2023 International Conference on Computer and Applications (ICCA)* (pp. 1-7). IEEE. https://doi.org/10.1109/icca59364.2023.10401723

AUTHOR:

Francisco Leslie López del Castillo Wilderbeek: PhD in communication from the Universitat Pompeu Fabra, Collaborating Professor at the UOC, Associate Professor at the UOC. He works as a media analyst at Rebold. His recent scientific production has focused on the origin and effects of Generative Artificial Intelligence (GenAI) in articles such as: "Generative Artificial Intelligence: technological determinism or socially constructed artifact" (*Palabra Clave*).
franciscoleslie@alumni.upf.edu

Orcid ID: https://orcid.org/0000-0002-6664-7849
Google Scholar: https://scholar.google.es/citations?user=UFyl1AkAAAAJ&hl=es&oi=ao

RELATED ARTICLES:

Abdullah, N. H., Hassan, I., Azura Tuan Zaki, T. S., Ahmad, M. F., Hassan, N. A., Mohd Zahari, A. S., Ismail, M. M., & Azmi, N. J. (2022). Examining the Relationship Between Factors Influencing Political Information Seeking-Behaviour through Social Media among Youths in Malaysia. *Revista de Comunicación de la SEECI*, 55, 1-15. https://doi.org/10.15198/seeci.2022.55.e746

Encinillas García, M., & Martín Sabarís, R. (2023). Desinformación y Salud en la era PRECOVID: Una revisión sistemática. *Revista de Comunicación y Salud*, 13, 1-15. https://doi.org/10.35669/rcys.2023.13.e312

Gómez-Diago, G. (2022). Perspectivas para abordar la inteligencia artificial en la enseñanza de periodismo. Una revisión de experiencias investigadoras y docentes. *Revista Latina de Comunicación Social*, 80, 29-46. https://doi.org/10.4185/RLCS-2022-1542

Martínez-Sánchez, J. A. (2022). Prevención de la difusión de fake news y bulos durante la pandemia de COVID-19 en España. De la penalización al impulso de la alfabetización informacional. *Revista de Ciencias de la Comunicación e Información*, 27, 15-32. https://doi.org/10.35742/rcci.2022.27.e236

Pinto, H., Nogueira, C., & Vieira, G. (2023). Digitalisation landscape in the European Union: Statistical insights for a Digital Transformation. *European Public & Social Innovation Review*, *8*(1), 20-38. https://pub.sinnergiak.org/esir/article/view/233