COMMUNICATION & SOCIETY

Special

Transparency in AI usage within fact-checking platforms in Spain and its ethical challenges

Roger Cuartielles
https://orcid.org/0000-0001-6226-6697
roger.cuartielles@upf.edu
Universitat Pompeu Fabra

Marcel Mauri-Ríos
https://orcid.org/0000-0003-2615-8343
marcel.mauri@upf.edu
Universitat Pompeu Fabra

Ruth Rodríguez-Martínez
https://orcid.org/0000-0001-5633-6126
ruth.rodriguez@upf.edu
Universitat Pompeu Fabra

Submitted
April 24th, 2024
Approved
July 5th, 2024

Communication & Society
ISSN 0214-0039
E ISSN 2386-7876
www.communication-society.com

2024 – Vol. 37(4)
pp. 257-271

How to cite this article:
Cuartielles, R., Mauri-Ríos, M. & Rodríguez-Martínez, R. (2024). Transparency in AI usage within fact-checking platforms in Spain and its ethical challenges, *Communication & Society, 37*(4), 257-271.
https://doi.org/10.15581/003.37.4.257-271

Abstract

Transparency –encompassing methodological, financial, and source-related aspects, as well as the tools employed– is central to the operations of professional fact-checking platforms. However, the growing adoption of artificial intelligence tools in fact-checking introduces new ethical challenges. This research investigates the extent to which these platforms believe they should disclose their use of AI and assesses the current practices on their websites regarding this technology. The study employs a qualitative methodology, including semi-structured interviews with professionals from accredited Spanish verification platforms and content analysis of these organizations' websites. The findings indicate that transparency in AI usage is widely regarded as an ethical imperative. Nevertheless, the application of this standard often becomes ambiguous when addressing specific practices and cases. Many professionals question the necessity of explicitly disclosing AI usage when the technology primarily supports the verification and is overseen by human reviewers. Additionally, a lack of understanding of AI's functionality sometimes hinders the ability to identify whether the tools employed incorporate AI. The content analysis also reveals that explicit mentions of AI use on the websites are rare and that platforms lack open-access manuals or protocols that outline and regulate their AI practices.

Keywords

Fact-checking, transparency, Artificial Intelligence, ethics, journalism.

Funding

This research was funded by the project titled "Accountability Instruments in the Face of Disinformation: Impact of Fact-Checking Platforms as Accountability Tools and Curriculum Proposal" (PID2019-106367GB-I00/AEI/10.13039/501100011033) (FACCTMedia). It is supported by the Ministry of Science and Innovation of Spain (2020-2024) in collaboration with the Political Communication, Journalism, and Democracy Research Group at Pompeu Fabra University (POLCOM-GRP) and the AGAUR Grants (2021 SGR 00486) from the Generalitat of Catalonia. The author, Roger Cuartielles, holds an FPU contract from the Ministry of Universities of the Government of Spain, under reference FPU22/03075.

1. Introduction

The integration of artificial intelligence (AI) into content gathering, production, and dissemination is rapidly becoming a standard in the news industry. Its application in data verification and fact-checking represents a significant advancement (Beckett & Yaseen, 2023). In recent years, numerous fact-checking platforms have incorporated AI into their verification routines, using tools such as bots for debunking (Arias-Jiménez *et al.*, 2023; Pasquetto *et al.*, 2022; Flores-Vivar, 2020) and machine learning systems for detecting claims and manipulated videos, and photographs. This trend underscores AI's potential as a crucial tool in combating the disruptive effects of disinformation (Rubin, 2022).

AI has thus emerged as a valuable resource for streamlining various aspects of fact-checking workflows. It facilitates the monitoring and identification of claims (detection), data collection for content verification (reporting), and exposure of false information (debunking) (Guo *et al.*, 2022). By addressing the rapid spread (Vosoughi *et al.*, 2018) and the increasing sophistication of disinformation in the digital realm, AI-driven technologies reduce the time required for detection and enhance the response capabilities of fact-checkers against disinformation.

The integration of AI into verification has led to the emergence of terms such as "computational fact-checking" and "automated fact-checking" (AFC) (Thorne & Vlachos, 2018), which refer to practices that automate parts of the fact-checking process using AI. However, it is crucial to emphasize that these processes remain only partially automated, as human oversight is fundamentally required for verification (Graves, 2018). Consequently, some researchers have coined the term "assisted fact-checking" to describe this human-in-the-loop process (La-Barbera *et al.*, 2022).

Rigor, impartiality, accountability, and transparency –in both methodological approaches and funding sources– are considered critical elements in the professional conduct of fact-checkers (Singer, 2021). Since its inception, fact-checking has prioritized the disclosure and accessibility of sources as a fundamental practice, recognizing that the credibility of fact-checking hinges on the demonstration of rigorous and accurate verification processes. This practice is essential, given the potential for skepticism surrounding the accuracy and integrity of these processes (Graves, 2016).

Transparency has thus become a hallmark of fact-checking platforms (Moreno-Gil *et al.*, 2022; Moreno-Gil *et al.*, 2021), particularly in methodological aspects. Professional fact-checkers publicly disclose their sources for each fact-check and the tools employed (Vizoso & Vázquez-Herrero, 2019). Dedicated sections on their websites often explain the methodologies used and the scales applied to measure the accuracy of information (López-Pan & Rodríguez-Rodríguez, 2020). These disclosures aim not only to enhance the reliability of the information but also to empower citizens by enabling them to trace and replicate verification (Graves, 2016).

In this context, the integration of AI in data verification must adhere to the same transparency principles that guide all fact-checkers' work. In recent years, platforms like the British organization *Full Fact* have introduced dedicated sections on their websites detailing the use of AI within their operations, including information on the team members responsible for developing this technology.

This research aims to: 1) examine the extent to which professionals at Spanish fact-checking platforms believe they should be transparent about their use of AI in their work; and 2) assess the level of transparency these organizations currently provide on their websites regarding the use of AI-supported technology.

This study builds on previous studies exploring the use and experimentation with AI by Spanish fact-checking platforms (Larraz *et al.*, 2023; Cuartielles *et al.*, 2023). However, there is a notable gap in research specifically addressing transparency in AI usage within the fact-checking context in Spain. Therefore, this study serves as a preliminary effort to bridge this gap.

ISSN 2386-7876 – © 2024 Communication & Society, 37(4), 257-271

2. Theoretical Framework

Transparency in journalism has traditionally been defined as the public's ability to trace, verify, critique, and even participate in the journalistic process (Deuze, 2005). It is also considered a fundamental ethical principle for journalists (Plaisance, 2007) and serves as a tool for audiences to assess the reliability of information (Kovach & Rosenstiel, 2007). Furthermore, transparency is seen as a remedy to the profession's ongoing credibility crisis (Singer, 2021).

Conceptually, transparency in journalism has evolved into a multifaceted requirement that includes the dissemination of public information. It involves two key aspects: open access to corporate data, including media principles, composition, and organizational structure; and the public explanation of editorial processes. Consequently, transparency has become a reliable indicator of media accountability alongside self-regulation and public participation (Ramon-Vegas & Mauri-Ríos, 2020).

Transparency has also emerged as a fundamental requirement in the use of AI within journalistic organizations (Diakopoulos & Koliska, 2017). This is particularly crucial given that decision-making processes and operations –such as prioritization, classification, association, and information filtering– can be delegated to algorithms, which are often perceived as opaque "black boxes" (Diakopoulos, 2015). In this regard, numerous studies have established that transparency has become a policy guideline in global documents and proposals for the ethical use of AI (Cotino-Hueso, 2022). Notable contributions include those by Jobin *et al.* (2019), Fjeld *et al.* (2020) and Hagendorff (2020), where transparency in AI use is closely linked to the concept of explainability.

According to the European Commission (2019), explainability in AI use "concerns the ability to explain both the technical processes of an AI system and the related human decisions" (p. 22). The European Commission (2019) also associates transparency with traceability and communication.

Traceability involves documenting, gathering, and labeling data and the algorithms employed. Communication refers to publicly identifying AI systems as non-human instruments, enabling users to recognize when they are interacting with such systems. It also includes disclosing the capabilities and limitations of AI systems through an appropriate information model that details the system's accuracy and limitations (European Commission, 2019).

Similarly, the UNESCO (2022) guidelines on AI ethics stress the importance of transparency and explainability. They emphasize fully informing the public whether a decision was based on or influenced by algorithms and recommend offering users the option to request explanations and information about implementation protocols. Thus, transparency and the intelligibility of AI are presented as essential for safeguarding human rights and ensuring individual and collective self-determination (Mantelero, 2022).

In journalism, organizations such as Reporters Without Borders (RSF) highlight transparency as an essential ethical requirement in AI usage. They assert, "Any use of AI that has a significant impact on the production or distribution of journalistic content should be clearly disclosed and communicated to everyone receiving information alongside the relevant content" (RSF, 2023, p. 2). Furthermore, RSF advocates for media organizations to maintain a public record of the AI systems they use, detailing their objectives, operational scope, and implementation conditions (RSF, 2023).

Leading media organizations, such as the BBC, have also established principles for using AI, identifying transparency as one of the nine core aspects guiding the implementation of this technology. Here, transparency is directly linked to explainability, defined as clearly identifying for the audience which content involves AI, the data collected for its implementation, and the reasons for using the technology and its potential impact on the audience (BBC, 2024).

Few organizations have published ethical guidelines regarding AI use in journalism in Spain. For instance, guidelines from the Federation of Associations of Journalists of Spain (FAPE)

ISSN 2386-7876 – © 2024 Communication & Society, 37(4), 257-271

lack specific references to AI. However, in Catalonia, the Consell de la Informació de Catalunya (CIC), endorsed by the Col·legi de Periodistes de Catalunya, has published a manual offering ethical recommendations for the use of algorithms in newsrooms.

In the manual published by the CIC, transparency, coupled with accountability, is emphasized as a fundamental ethical requirement in the use of AI. This requirement applies to the management of data used to train and implement AI technologies and to the algorithmic processes and outcomes they produce (Ventura Pocino, 2021). Additionally, the EFE Agency's style guide incorporates references to AI management, stating that it is unnecessary to disclose the use of AI tools if they are not directly generating content and remain under human supervision (EFE, 2024). This approach contrasts somewhat with the CIC guidelines, which advocate for communicating AI usage whenever possible.

Fact-checking, which emerged from within the media system and is closely aligned with journalism through its core practice of data verification (Redondo, 2018), is similarly subject to transparency guidelines. While no ethical guidelines specifically address AI use in fact-checking, organizations such as the International Fact-Checking Network (IFCN) have long emphasized transparency in their code of principles. The IFCN requires that fact-checking organizations publicly disclose the methodologies used for selecting, investigating, writing, and publishing verifications (IFCN, 2020). Similarly, the European Fact-Checking Standards Network (EFCSN) code of good practices mandates transparency, particularly in publicly disclosing each organization's composition, structure, and funding sources (EFCSN, 2022).

Collaborative partnerships between fact-checking agencies and technology platforms like Meta have also been noted by researchers such as Bélair-Gagnon *et al.* (2023). These collaborations aim to secure funding, increase online traffic, enhance visibility, and provide technological training. However, critics like Castellet *et al.* (2023) and Rúas-Araújo & Fontenla-Pedreira (2024) caution that these collaborations could lead to problematic economic dependencies, as they involve major internet and social media companies whose business models do not necessarily prioritize the pursuit of truth.

Research on AI applications in fact-checking is still in its nascent stages. However, several Spanish fact-checking organizations have begun integrating AI into their operations. For example, *Maldita.es* and *EFE Verifica* use chatbots via WhatsApp to handle verification requests, while *Newtral* has implemented tools like ClaimHunter to identify political claims on the X platform and Editor to analyze audiovisual content. The Editor tool uses automatic transcription and Microsoft's BERT language model to flag factual statements that need checking. These statements are then reviewed via a Slack channel where a team of experts examines them to decide if further verification is needed before publication (Cuartielles & Carral, In press; Cuartielles *et al.*, 2023).

Moreover, *Newtral*, in collaboration with ABC (Australia), has developed an AI program called ClaimCheck, which aims to detect repeated false information in political discourse through a semantic similarity model (Larraz *et al.*, 2023). Other Spanish fact-checking platforms utilize InVID, which has been enhanced with AI to offer improved tools for video frame and image searches, as well as optical character recognition (OCR) for detecting image manipulation. Additionally, *EFE Verifica* employs Remini.ai to enhance image clarity and highlight key objects in verification processes. Both *EFE Verifica* and *VerificaRTVE* use detection, archiving and automatic transcription tools from the IVERES (Identification, VERification, and RESponse) initiative, an ongoing project funded by the Ministry of Science and Innovation that involves universities and fact-checkers in developing an AI toolkit for detecting false information (Cuartielles & Carral, In press). Furthermore, academic initiatives like the DEBATrue project are exploring AI and blockchain applications in verification, collaborating with media verification teams from RTVE and Agencia EFE (Pérez-Curiel *et al.*, 2023).

ISSN 2386-7876 – © 2024 Communication & Society, 37(4), 257-271

3. Methodology

This study aimed to investigate the extent to which professionals at Spanish fact-checking platforms perceive the need for transparency in their use of AI and to examine the level of transparency these organizations provide on their websites regarding this technology.

The analysis focused on six Spanish fact-checking platforms listed as active in the Duke Reporters' Lab database: *Maldita.es*, *Newtral*, *EFE Verifica*, *AFP Factual España*, *Verificat,* and *Infoveritas*. These organizations are also signatories of the International Fact-Checking Network (IFCN) Code of Principles, which emphasizes five core principles: 1) non-partisanship and fairness; 2) transparency regarding sources; 3) transparency of funding, and organization; 4) standards and transparency of methodology; and 5) an open and honest corrections policy. To enhance representativeness, *VerificaRTVE*, a Spanish platform not listed in the IFCN Code of Principles or the Duke Reporters' Lab database but registered with the European Digital Media Observatory (EDMO) as a fact-checking entity, was also included.

Table 1. Characteristics of the fact-checking platforms participating in the study.

Platform	Website	Creation	No of Fact-checker employees
Maldita.es	*https://maldita.es*	2018	24
Newtral	*https://www.newtral.es/*	2018	13
EFE Verifica	*https://verifica.efe.com/*	2019	9
Verificat	*https://www.verificat.cat/*	2019	6
AFP Factual España	*https://factual.afp.com/afp-espana*	2019	2
VerificaRTVE	*https://www.rtve.es/noticias/verificartve/*	2020	5
Infoveritas	*https://info-veritas.com/*	2021	5

Source: Own elaboration based on interviews with fact-checking platforms.

To address the first objective of this research, seven semi-structured interviews were conducted with senior professionals in the fact-checking field, including editors and section heads: Pablo Hernández (Academic Research Coordinator at *Maldita.es*); Irene Larraz (Coordinator at *Newtral Lab*); Sergio Hernández (Head of *EFE Verifica*); Alba Tobella (Co-founder and Head of Content at *Verificat*); Natalia Sanguino (Editor at *AFP Factual España*); Blanca Bayo (Deputy Head of *VerificaRTVE*); and Guillermo García (Editor-in-Chief at *Infoveritas*).

The interviews, lasting between 20 and 60 minutes, were conducted in March 2024 via Google Meet to accommodate the geographical dispersion of the participants across Spain. All interviews were recorded and subsequently transcribed for analysis. The transcription was carried out using Trint, an AI-driven tool for automatic transcription. The research team manually reviewed each transcript to ensure data accuracy. The qualitative analysis software ATLAS.TI was employed during the coding phase, and the constant comparative method was applied (Wimmer & Dominick, 2013). The data were categorized, and relationships and themes were refined through an iterative process to identify recurring issues. The semi-structured interview format offered flexibility, with questions organized into two main thematic blocks: a) the use of AI on the platform and b) perceptions of transparency regarding AI use.

To achieve the second research objective, a qualitative web content analysis (Herring, 2010) was conducted in March 2024. This analysis aimed to assess whether the organizations that acknowledged AI use in their work routines during the interviews adhered to specific indicators of transparency and public explainability on their websites. A custom codebook was developed, drawing on key elements identified by authoritative bodies such as the European Commission

ISSN 2386-7876 – © 2024 Communication & Society, 37(4), 257-271

(2019), UNESCO (2022), Reporters Without Borders (RSF, 2023), and the BBC (2024) as essential for the ethical use of AI technology. The codebook included seven categories relevant to AI transparency in fact-checking, such as explicit disclosure of AI tools used in published fact-checks, informative pieces on AI adoption, a dedicated AI section on the website, and references to AI tools in the methodology section. These categories provided a structured approach to evaluate the level of transparency offered by the selected fact-checking platforms.

Table 2. Codebook.

Explicit mention of the use of AI in published fact-checks in case of using tools that apply this technology.
Production of informative pieces about the adoption of AI tools within the organization.
Mention of AI tool usage in the website's methodology section.
Existence of a dedicated website section informing about AI usage.
Explicit mention of AI use when employing a chatbot for user inquiries.
Identification on the platform's website of team members responsible for developing or adopting AI.
Existence of an open-access protocol or manual on AI use on the platform's website.

Source: Own elaboration based on the broad recommendations of the European Commission (2019), UNESCO (2022), RSF (2023), and the BBC (2024).

4. Results

The findings from the interviews with the fact-checking professionals reveal that all Spanish fact-checking platforms, except for Verificat, utilize AI-powered tools. Interviewees unanimously viewed transparency in AI use as "necessary," with many advocating for it to be "absolute" or "as comprehensive as possible." Despite this consensus, public disclosure of AI usage across these platforms is limited. Among the platforms surveyed, only *Newtral* explicitly mentions the use of AI in its methodology section. This study also uncovers varying interpretations of transparency implications in AI usage for fact-checking. Some professionals expressed uncertainty about the relevance of disclosing AI use when the technology merely supports verification processes described as "non-automated," emphasizing that human judgment remains central to supervision and verification.

The results are presented in three distinct sections. First, the study details the current use of AI among Spanish fact-checking platforms. Next, it explores transparency issues, focusing on Spanish fact-checking professionals' perceptions. Lastly, based on content analysis, the study specifies the level of transparency demonstrated by these platforms regarding their AI use.

4.1. *Use of AI*

AI usage in fact-checking in Spain dates back to 2018, when *Newtral* began employing AI-driven systems for monitoring misinformation, making it the earliest instance of such technology in the field. Following *Newtral*, *Maldita.es*, *EFE Verifica*, *VerificaRTVE*, and *AFP Factual España* adopted AI between 2019 and 2021, with many of these implementations accelerated by the challenges posed by the COVID-19 pandemic. *Infoveritas*, the most recently established platform, introduced AI tools between 2021 and 2022.

Most of these platforms utilize AI primarily for detecting misinformation, monitoring content across media outlets and social networks –a practice often referred to as "social listening"– and identifying specific manipulations within audiovisual content. The use of chatbots on platforms like WhatsApp serves a dual purpose: they help detect disinformation through user inquiries while also serving as a distribution channel for fact-checks. These chatbots, powered by natural language processing (NLP), match citizen inquiries with previously conducted verifications, thereby facilitating the efficient dissemination of fact-checks. Table 3 categorizes the main AI tools identified by the study participants into proprietary and external tools.

ISSN 2386-7876 – © 2024 Communication & Society, 37(4), 257-271

Table 3. The main AI tools used by Spanish fact-checking platforms categorized into proprietary and external tools.

Platform	Proprietary AI tools	External AI tools
Newtral	▪ ClaimHunter (*claims identification on X*). ▪ Editor (*claims identification in audiovisual statements*). ▪ ClaimCheck (*detection of repeated misinformation*).	▪ InVID (*plugin with AI detection tools such as forensic image analysis*). ▪ PimEyes (*facial recognition*). ▪ Chatbot.
Maldita.es	▪ Chatbot. ▪ Propriety tool for detecting disinformation narratives among alerts received through the chatbot.	▪ InVID. ▪ Hive (*detection of AI-generated content)*.
VerificaRTVE	▪ Detection, archiving and automatic transcription tools developed by IVERES (*RTVE's proprietary project*).	▪ InVID. ▪ Trint (*automatic transcription*). ▪ AI or Not (*detection of AI-generated content*). ▪ Hive.
EFE Verifica	▪ No publicly disclosed proprietary tools.	▪ InVID. ▪ Remini.ai (*image enhancement)*. ▪ Meltwater *(identification of viral content).* ▪ IVERES tools. ▪ Chatbot.
AFP Factual España	▪ InVID.	▪ Chatbot.
Infoveritas	▪ Tool using NLP to detect hoaxes on social media.	▪ Hugging Face *(detection of AI-generated content)*. ▪ Hive.

Source: Own elaboration based on data provided by the interviewees.

4.2. *Perceptions of transparency in AI usage*

The fact-checking professionals interviewed unanimously consider transparency essential for maintaining public trust in their work. Guillermo García from *Infoveritas* emphasizes, "[...] it's crucial to explain the process step by step so that people understand how it works –not just in a methodology section, but by clearly demonstrating each step" (Personal communication, March 7, 2024). In this context, transparency serves a dual purpose: ensuring accountability and educating the public on verification processes, thereby empowering citizens to counter misinformation. Similarly, Blanca Bayo from *VerificaRTVE* highlights the educational aspect of transparency:

> [...] the goal is also to educate people to recognize when content is false and follow relatively simple steps to develop their own tools for distinguishing what's true from what isn't. We always specify what we use, and if something doesn't work out, we also disclose that (Personal communication, March 6, 2024).

Regarding AI-powered tools, the fact-checkers generally agree that transparency in their use should be a central focus. Sergio Hernández from *EFE Verifica* notes, "[...] transparency should be present throughout the entire fact-checking process, and even more so" (Personal communication, March 1, 2024). Natalia Sanguino from *AFP Factual España* further emphasizes that transparency should be "as comprehensive as possible, with prior reflection," and directed toward two audiences: the public, and the tool developers to facilitate continuous improvement (Personal communication, March 7, 2024).

However, when discussing AI's role in fact-checking, the interviews revealed a more nuanced understanding of transparency. Transparency in fact-checking generally centers on

ISSN 2386-7876 – © 2024 Communication & Society, 37(4), 257-271

methodological aspects, such as describing the steps professionals take to verify content and the sources consulted. As Irene Larraz from *Newtral* explains, "We're transparent about everything; it's central to our philosophy. We disclose our sources, data, tools, and methodology" (Personal communication, March *5, 2024).* Financial transparency is also emphasized, with Guillermo García from *Infoveritas* stating, "It's crucial to highlight all matters related to funding, including how we finance ourselves and who is behind it."

Nevertheless, public disclosure of the specific use of AI within individual tools is often considered a secondary concern. The consensus among the professionals is that if the verification process is clearly outlined and the tools used are mentioned, specifying whether these tools incorporate AI is not deemed crucial. This detail is regarded as an "internal functionality" issue, with the belief that transparency is sufficiently achieved by acknowledging the tool itself. Natalia Sanguino from *AFP Factual España* encapsulates this view: "If the verification is clear and the steps are straightforward, that suffices. Ultimately, AI is just a technology within a specific tool." Similarly, Blanca Bayo from *VerificaRTVE* argues, "Since we don't generate content with artificial intelligence, there's no need to specify further, other than citing all our sources and the tools we use."

One key aspect that emerged from the fact-checkers' reflections on the need to publicly disclose the use of AI was the impact this technology's application might have on users. Professionals like Pablo Hernández from *Maldita.es* emphasize the importance of first assessing the extent of human involvement in each process to determine whether it is necessary to publicly specify AI's role:

> In our use of AI for detecting narratives, which doesn't directly impact users as it merely provides clues on where to direct our work, it's not essential to disclose it. The product the user consumes is entirely controlled by the *Maldita* team, which publishes all the articles. It's not a case of "This was done by AI without human oversight," but rather "This was done by a team of humans guided by AI to direct their work" (Personal communication, March 1, 2024).

Irene Larraz from *Newtral* echoes this sentiment, stressing the auxiliary role of AI alongside the core human element in fact-checking:

> [...] all our tools are always hybrid. There is always a human in the loop. A fact-checker supervises the work done by the tool. The process is never fully automated, not even during the detection phase. AI provides leads; we do all the actual work ourselves.

Larraz compares AI's role in fact-checking to other supportive tools like Microsoft Word, noting that such tools are integral to daily operations but are not explicitly mentioned in reports because of their non-determinative role in the verification process. Bayo from *VerificaRTVE* similarly argues that specifying AI use in particular tools is unnecessary and highlights the difficulty in identifying which tools employ AI, complicating efforts to maintain transparency:

> We're not engineers, so we don´t specify how AI is used because we aren't entirely sure ourselves. We´ve taken several courses and have access to tools we can utilize. We use them but don't go into detail about their workings because we don't fully understand them.

This uncertainty among fact-checkers about whether specific processes involve AI is evident. For instance, Pablo Hernández from *Maldita.es* questions whether reverse image searches utilize this technology, while Alba Tobella from *Verificat* is unsure if automated systems that match citizen queries with previously conducted verifications involve AI: "To me, this feels like technology, but I'm not sure if it qualifies as AI; for me, AI is about creating content." Meanwhile, professionals like Bayo from *VerificaRTVE* stress that: "AI is embedded in many of the processes we take for granted." Some fact-checkers, including Natalia Sanguino from *AFP Factual España*, argue that "there is a need for much more educational guidance on the use of AI," noting that "there's considerable confusion" among verification professionals and the public about AI.

ISSN 2386-7876 – © 2024 Communication & Society, 37(4), 257-271

These discussions highlight the ethical dilemmas introduced by AI in fact-checking, particularly regarding transparency about its application. As Sergio Hernández from *EFE Verifica* points out, "Completely infallible AI tools have yet to be developed," necessitating caution when implementing them. Additionally, using specific AI tools to detect AI-generated content presents new methodological challenges. Pablo Hernández from *Maldita.es* expressed concerns over AI tools that use percentages to detect manipulated content, complicating the determination of whether content is "entirely false." He explains, "We can't be 80% or 99% certain that something is false information due to these limitations." This uncertainty may require fact-checkers to rethink their evaluation criteria.

The interviews also revealed that the fact-checkers have not received explicit guidelines from the two main organizations that oversee most of the world's fact-checking platforms–the International Fact-Checking Network (IFCN) and the European Fact-Checking Standards Network (EFCSN)–regarding the level of transparency required in their use of AI. Nonetheless, professionals like Sergio Hernández from *EFE Verifica* acknowledge that this issue is a topic of ongoing debate within the fact-checking community. He noted that organizations like the EFCSN are considering implementing standardized disclosure guidelines on AI usage to "reach an agreement on best practices."

4.3. *Level of transparency in the use of AI*

The qualitative content analysis reveals that explicit mentions of AI tools in published fact-checks are rare. For instance, platforms such as *Newtral* do not usually specify whether AI tools are used for monitoring and detecting misinformation in its public corrections. Similarly, *Infoveritas* does not disclose its use of NLP tools to detect hoaxes on social media, nor does *AFP Factual España* mention the use of AI when using InVID in its verifications. Other platforms tend to make token mentions of AI-powered tools only in their debunking reports.

However, we observed that organizations such as *Maldita.es*, *VerificaRTVE*, *EFE Verifica*, and *Infoveritas* are more likely to specify their use of AI when employing tools specifically designed to identify AI-generated content. Examples include Hive, Remini.ai, and Hugging Face.

Newtral distinguishes itself by regularly publishing informative pieces about its use of AI, particularly regarding the development and adoption of new tools, as it allocates the most resources to AI implementation. In contrast, *AFP Factual España* and *VerificaRTVE* produce far fewer informative pieces, even regarding their proprietary tools like InVID or the IVERES Project tools, respectively. Most of their informative content is disseminated through their respective parent organizations, AFP and RTVE, respectively, rather than through their own websites. Notably, none of the pieces from AFP on the development of InVID explicitly mention the use of AI. Meanwhile, *Maldita.es* and *EFE Verifica* have published pieces disclosing their use of a chatbot for receiving citizen inquiries to "automatically verify hoaxes," although they did not explicitly reference AI. Similarly, *Infoveritas* has published content describing its use of "cutting-edge technology" without clarifying if this technology involves AI.

Newtral is the only fact-checking platform in the sample that mentions the use of AI tools in its website's methodology section and posts pieces about adopting and developing proprietary tools. The other platforms do not mention the use of AI tools in this section, whether proprietary or external. Moreover, organizations like *VerificaRTVE* stand out for not having a methodology section at all.

The use of WhatsApp chatbots for user interaction is one of the most widely employed AI tools by most fact-checking platforms in Spain. However, none of the four organizations using these chatbots (*Maldita.es*, *EFE Verifica*, *Newtral,* and *AFP Factual España)* explicitly clarify that the tool is AI-supported. While terms implying "automated processes and operations" are used, none specifically mention the use of AI in the context of chatbot interactions. Addressing this issue in the qualitative interviews, professionals such as Sergio Hernández from *EFE Verifica* explained that they deliberately avoided using the term AI because they considered it a

ISSN 2386-7876 – © 2024 Communication & Society, 37(4), 257-271

"technicality" that might create reservations among users. Instead, they chose to present the WhatsApp channel as "a service with some automated responses," emphasizing that "most of the content users interact with is created by platform professionals."

Furthermore, none of the Spanish fact-checking platforms that use AI tools in some processes feature a dedicated section on their websites explaining their use of this technology. However, they do mention whether they have engineers in their respective team sections, as seen with *Newtral*, *Maldita.es*, *AFP Factual España* and *Infoveritas* –all of which employ such professionals. Nevertheless, *Newtral* and *Infoveritas* are the only organizations where the involvement of specific engineers in AI tool development and adoption is clearly identifiable.

None of the Spanish fact-checking platforms utilizing AI tools currently provide an open-access protocol or user manual for this technology on their websites. As noted during the qualitative interviews, some progress has been made in this area. *Newtral*, for instance, is developing a publicly accessible protocol, *AFP Factual España* follows an internal protocol created by AFP regarding the general use of AI, and *EFE Verifica* adheres to its parent agency's directives, which have recently incorporated references to AI usage in its style guide. In contrast, *Maldita.es* and *Infoveritas* lack publicly accessible manuals or protocols on AI use, although both acknowledge having had internal discussions on the matter.

Based on these considerations, Table 4 illustrates the level of compliance with each analyzed transparency indicator among the fact-checking platforms. The table uses color-coding to represent different levels of compliance. Green indicates the highest degree of compliance, where aspects are explained in detail and frequently. Orange signifies medium compliance, where aspects are mentioned occasionally. Light orange represents the lowest degree of compliance, where aspects are barely mentioned. Gray cells indicate when a platform does not meet the indicator due to the absence of the relevant tool.

Table 4. Level of compliance with the analyzed transparency indicators.

Indicators	*Newtral*	*Maldita.es*	*AFP Factual España*	*VerificaRTVE*	*EFE Verifica*	*Infoveritas*
Explicit mention of the use of AI in published fact-checks in case of using tools that apply this technology[1].						
Production of informative pieces about the adoption of AI tools in the organization.						
Mention of AI tool usage in the website's methodology section.						
Explicit mention of AI usage when employing a chatbot for user inquiries.						

[1] Verifications published in the last two years were considered.

ISSN 2386-7876 – © 2024 Communication & Society, 37(4), 257-271

Existence of a dedicated section informing about AI usage.						
Identification on the platform's website of team members responsible for developing or adopting AI.						
Existence of an open-access protocol or user manual on AI usage on the platform's website.						

Source: Own elaboration.

5. Discussion and conclusions

Transparency in the use of AI within fact-checking is perceived as crucial, especially as verification professionals view it as an ethical benchmark. However, the scope and implications of this standard become blurred when it comes to its implementation and application in specific practices, as evidenced by this study's findings. Verification professionals recognize that transparency is strongly associated with methodological and financial aspects –dimensions that have traditionally been the focal points of ethical actions by fact-checkers to avoid suspicion (Graves, 2016). Consequently, these dimensions have been integrated into the codes of good practice of the two major organizations –the IFCN and the EFCSN– overseeing most of the world's fact-checking platforms.

The perception of AI as a supportive tool in fact-checking has led most platforms to view it as "non-essential" to provide detailed explanations each time AI is used, whether through proprietary or external tools, since the core activities of fact-checking rely fundamentally on human judgment. This perspective reaffirms the human-in-the-loop nature of fact-checking (La-Barbera *et al.*, 2022), reflecting the transparency guidelines outlined in Agencia EFE's 2024 style guide, which stipulates that detailed AI disclosure is not necessary if the tools are employed in a supportive capacity and under human supervision.

Our findings align with previous research on information professionals' perceptions of AI integration into their work routines. These studies often depict AI as a supportive resource that relieves professionals from repetitive tasks, thereby increasing their capacity for more complex, creative endeavors. Moreover, these studies emphasize the importance of human oversight and decision-making, reinforcing that human judgment remains paramount in supervising AI tools (Noain-Sánchez, 2022).

These findings further highlight the role of fact-checkers as contextual agents in response to the rise of AI (Cuartielles *et al.*, 2023). However, they also underscore the need for more comprehensive training in AI, as verification professionals generally demonstrate a limited understanding of the technology, often reducing it to its generative capabilities. A deeper understanding of AI would enable fact-checkers to make more informed use of these tools and better assess their advantages and limitations in the fact-checking process. This enhanced understanding could also improve transparency in AI usage, as current transparency levels fall short of recommendations from organizations like the European Commission (2019), UNESCO (2022), RSF (2023), and the BBC (2024), particularly regarding explainability and communication about tool usage. However, it is also true that, as Cotino-Hueso (2022) specifies, the degree of

ISSN 2386-7876 – © 2024 Communication & Society, 37(4), 257-271

transparency should be determined by the impact or potential harm of an AI system on rights and interests, as stipulated by the AI Act passed by the European Parliament during the course of this research.

The growing adoption of AI tools in professional fact-checking organizations has prompted a reconsideration of the transparency standards that have defined the fact-checking movement since its inception (Singer, 2021) and shaped its identity as a reformative movement (Graves & Cherubini, 2016). By defining and agreeing upon the level of transparency required in AI usage, the professional fact-checking community could strengthen its rigor in an increasingly deceptive information ecosystem, which could be further refined and accelerated through the use of AI by bad actors (Franganillo, 2022). This approach could leverage and highlight the potential of AI to uncover deceptions (Rubin, 2022). Enhancing communication and explainability in AI usage could further empower citizens, a function traditionally associated with the fact-checking phenomenon (Graves, 2016).

Incorporating recommendations on transparency and AI usage into the principal codes of good practice within the fact-checking field will be crucial for solidifying and standardizing the ethical application of this technology in verification work. Moreover, adopting more explanatory formats (Moreno-Gil *et al.*, 2023) will help communicate the implications of AI usage and strengthen the fight against the malicious uses of such technology.

While this research covers all accredited Spanish fact-checking platforms, it is important to note that the study is limited to the Spanish context and a specific period, which is subject to constant technological and regulatory changes. Nonetheless, this study opens a new avenue of research, specifically transparency in the application of AI within the field of informational verification.

References

Agencia EFE (2024). *Nuevo libro del estilo urgente*. Retrieved from https://recursos.efe.com/objetos_app/libroestilo/libroDelEstiloUrgente.pdf

Arias-Jiménez, B., Rodríguez-Hidalgo, C., Mier-Sanmartín, C. & Coronel-Salas, G. (2023). Use of chatbots for news verification. In P. C. López-López, D. Barredo, A. Torres-Toukoumidis, A. De-Santis & O. Avilés (Eds.), *Communication and applied technologies. Smart innovation, systems and technologies*, vol. 318 (pp. 133-143). Singapur: Springer Nature. https://doi.org/10.1007/978-981-19-6347-6_12

Beckett, C. & Yaseen, M. (2023). *Generating Change. A global survey of what news organisations are doing with AI.* The London School of Economics and Political Science, Google News Initiative. Retrieved from https://www.journalismai.info/research/2023-generating-change

Bélair-Gagnon, V., Larsen, R., Graves, L. & Westlund, O. (2023). Knowledge Work in Platform Fact-Checking Partnerships. *International Journal of Communication, 17*, 1169-1189. Retrieved from https://ijoc.org/index.php/ijoc/article/view/19851/4044

British Broadcasting Corporation (BBC) (2024, February 7). *BBC AI Principles*. Retrieved from https://www.bbc.co.uk/supplying/working-with-us/ai-principles/

Castellet, A., Varona, D. & Álvarez García, S. (2023). Capítulo 6. Verificadores en España: una visión de su lógica de negocio. *Espejo De Monografías De Comunicación Social, 13*, 119-136. https://doi.org/10.52495/c6.emcs.13.p99

Cotino-Hueso, L. (2022). Transparencia y explicabilidad de la inteligencia artificial y "compañía" (comunicación, interpretabilidad, inteligibilidad, auditabilidad, testabilidad, comprobabilidad, simulabilidad...). Para qué, para quién y cuánta. In L. Cotino-Hueso & J. Castellanos-Claramunt (Eds.), *Transparencia y explicabilidad de la inteligencia artificial* (pp. 25-67). Valencia: Tirant Lo Blanch Monografías. Retrieved from https://www.uv.es/cotino/publicaciones/libroabiertotp22.pdf

Cuartielles, R. & Carral, U. (In Press). Herramientas de IA contra la desinformación populista. In F. Guerrero-Solé & L. Pérez-Altable (Eds.), *La democracia en riesgo. ¿Internet e IA al servicio de los populismos?* Barcelona: UOC.

Cuartielles, R., Ramon-Vegas, X. & Pont-Sorribes, C. (2023). Retraining fact-checkers: The emergence of ChatGPT in information verification. *Profesional de la información, 32*(5), e320515. https://doi.org/10.3145/epi.2023.sep.15

Deuze, M. (2005). What is Journalism?: Professional identity and ideology of journalists reconsidered. *Journalism, 6*(4), 442-464. https://doi.org/10.1177/1464884905056815

Diakopoulos, N. & Koliska, M. (2017). Algorithmic Transparency in the News Media. *Digital Journalism, 5*(7), 809-828. https://doi.org/10.1080/21670811.2016.1208053

Diakopoulos, N. (2015). Algorithmic Accountability: Journalistic investigation of computational power structures. *Digital Journalism, 3*(3), 398-415. https://doi.org/10.1080/21670811.2014.976411

European Commission, High-Level Expert Group on AI (2019, April 8). *Ethics Guidelines for Trustworthy AI*. Retrieved from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

European Fact-Checking Standards Network (2022, Summer). *Code of Standards*. Retrieved from https://efcsn.com/code-of-standards/

Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. & Srikumar, M. (2020*). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*. Berkman Klein Center for Internet & Society at Harvard University. Retrieved from https://dash.harvard.edu/handle/1/42160420

Flores-Vivar, J. M. (2020). Datos masivos, algoritmización y nuevos medios frente a desinformación y *fake news. Bots* para minimizar el impacto en las organizaciones. *Comunicación y hombre, 16*, 101-114. https://doi.org/10.32466/eufv-cyh.2020.16.601.101-114

Franganillo, J. (2022). Contenido generado por inteligencia artificial: oportunidades y amenazas. *Anuario ThinkEPI, 16*. https://doi.org/10.3145/thinkepi.2022.e16a24

Graves, L. (2018). *Understanding the promise and limits of automated fact-checking*. Reuters Institute, University of Oxford. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/graves_factsheet_180226%20FINAL.pdf

Graves, L. (2016). *Deciding what's True. The rise of political fact-checking in American journalism*. New York: Columbia University Press.

Graves, L. & Cherubini, F. (2016). *The rise of fact-checking sites in Europe*. Reuters Institute for the Study of Journalism. Retrieved from https://reutersinstitute.politics.ox.ac.uk/our-research/rise-fact-checking-sites-europe

Guo, Z., Schlichtkrull, M. & Vlachos, A. (2022). A survey on automated fact-checking. *Transactions of the association for computational linguistics, 10*, 178-206. https://doi.org/10.1162/tacl_a_00454

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and machines, 30*(1), 99-120. https://doi.org/10.1007/s11023-020-09517-8

Herring, S. C. (2010). Web content analysis: Expanding the paradigm. In J. Hunsinger, L. Klastrup & M. Allen (Eds.), *International handbook of Internet research* (pp. 233-249). Dordrecht/ Heidelberg/London/New York: Springer. https://doi.org/10.1007/978-1-4020-9789-8

International Fact-Checking Network (2020, April). *Code of Principles*. Retrieved from https://ifcncodeofprinciples.poynter.org/know-more/the-commitments-of-the-code-of-principles

Jobin, A., Ienca, M. & Vayena, E. (2019). The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence, 1*, 389-399. https://doi.org/10.1038/s42256-019-0088-2

Kovach, B. & Rosenstiel, T. (2007). *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect, Completely Updated and Revised*. New York: Three Rivers Press.

ISSN 2386-7876 – © 2024 Communication & Society, 37(4), 257-271

La-Barbera, D., Roitero, K. & Mizzaro, S. (2022). A hybrid human-in-the-loop framework for fact checking. In *Proceedings of the 6th Workshop on natural language for artificial intelligence* (NL4AI 2022), 3287. Retrieved from https://ceur-ws.org/Vol-3287/paper4.pdf

Larraz, I., Míguez, R. & Sallicati, F. (2023). Semantic similarity models for automated fact-checking: ClaimCheck as a claim matching tool. *Profesional de la Información, 32*(3), e320321. https://doi.org/10.3145/epi.2023.may.21

López-Pan, F. & Rodríguez-Rodríguez, J. M. (2020). El *fact-checking* en España. Plataformas, prácticas y rasgos distintivos. *Estudios sobre el mensaje periodístico, 26*(3), 1045-1065. https://doi.org/10.5209/esmp.65246

Mantelero, A. (2022). *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*. Berlin: Springer, Information Technology and Law Series.

Moreno-Gil, V., Ramon-Vegas, X., Rodríguez-Martínez, R. & Mauri-Ríos, M. (2023). Explanatory Journalism within European Fact Checking Platforms: An Ally against Disinformation in the Post-COVID-19 Era. *Societies, 13*, 237. https://doi.org/10.3390/soc13110237

Moreno-Gil, V., Ramon-Vegas, X. & Mauri-Ríos, M. (2022). Bringing journalism back to its roots: examining fact-checking practices, methods, and challenges in the Mediterranean context. *Profesional de la Información, 31*(2), e310215. https://doi.org/10.3145/epi.2022.mar.15

Moreno-Gil, V., Ramon-Vegas, X. & Rodríguez-Martínez, R. (2021). Fact-checking interventions as counteroffensives to disinformation growth: Standards, values, and practices in Latin America and Spain. *Media and Communication, 9*(1), 251-263. https://doi.org/10.17645/mac.v9i1.3443

Noain-Sánchez, A. (2022). Addressing the Impact of Artificial Intelligence on Journalism: the perception of experts, journalists and academics. *Communication & Society, 35*(3), 105-121. https://doi.org/10.15581/003.35.3.105-121

Pasquetto, I. V., Jahani, E., Atreja, S. & Baum, M. (2022). Social debunking of misinformation on WhatsApp: the case for strong and in-group ties. In *Proceedings of the ACM on human-computer interaction*, *6* (pp. 1-35). https://doi.org/10.1145/3512964

Pérez-Curiel, C., Rúas-Araújo, J. & Paiagua-Rojano, F. J. (2023). Desinformación y verificación de noticias políticas en las aulas: DEBATrue como aplicación digital para la educación mediática. In C. Hervás-Gómez, P. Román Graván, J. García Jiménez & C. Argüello Gutiérrez (Coords.), *Conexiones digitales: las tecnologías como puentes de aprendizaje* (pp. 67-83). Madrid: Dykinson.

Plaisance, P. L. (2007). Transparency: An Assessment of the Kantian Roots of a Key Element in Media Ethics Practice. *Journal of Mass Media Ethics, 22*(2–3), 187-207. https://doi.org/10.1080/08900520701315855

Ramon-Vegas, X. & Mauri-Ríos, M. (2020). Participación de la audiencia en la rendición de cuentas de los medios de comunicación: instrumentos de *accountability* y su percepción por parte de los ciudadanos españoles (Audience participation for media accountability: instruments and their perception by Spanish citizens). *RAEIC, Revista de la Asociación Española de Investigación de la Comunicación, 7*(13), 50-76 https://doi.org/10.24137/raeic.7.13.3

Redondo, M. (2018). *Verificación digital para periodistas. Manual contra bulos y desinformación internacional*. Barcelona: UOC.

Reporters sans frontières, RSF (2023, November 10). *Paris Charter on AI and Journalism*. Retrieved from https://rsf.org/sites/default/files/medias/file/2023/11/Paris%20Charter%20on%20AI%20and%20Journalism.pdf

Rúas-Araújo, J. & Fontenla-Pedreira, J. (2024). Contra la desinformación en red: la necesidad de una mirada crítica y enfoque multidisciplinar [Against online disinformation: the need for a critical look and multidisciplinary approach]. *Infonomy, 2*(2) e24024. https://doi.org/10.3145/infonomy.24.024

Rubin, V. L. (2022). *Misinformation and Disinformation: Detecting Fakes with the Eye and AI*. Cham: Springer.

Singer, J. B. (2021). Border patrol: the rise and role of fact-checkers and their challenge to journalists' normative boundaries. *Journalism, 22*(8), 1929-1946. https://doi.org/10.1177/1464884920933137

Thorne, J. & Vlachos, A. (2018). Automated fact checking: task formulations, methods and future directions. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 3346-3359). Retrieved from https://aclanthology.org/C18-1283

UNESCO (2022). *Recommendation on the Ethics of Artificial Intelligence*. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000381137

Ventura Pocino, P. (2021). Algoritmes a les redaccions: Reptes i recomanacions per dotar la intel·ligència artificial dels valors ètics del periodisme. Consell de la Informació de Catalunya. Retrieved from https://fcic.periodistes.cat/wp-content/uploads/2022/02/algorismes_a_les_redaccions_CAT_.pdf

Vizoso, A. & Vázquez-Herrero, J. (2019). Fact-checking platforms in Spanish. Features, organisation and method. *Communication & Society, 32*(1), 127-142. https://doi.org/10.15581/003.32.37819

Vosoughi, S., Roy, D. & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146-1151. https://doi.org/10.1126/science.aap9559

Wimmer, R. D. & Dominick, J. R. (2013). *Mass media research: an introduction*. Belmont, CA: Cengage Learning.