VISUAL REVIEW | Vol. 18, No. 2, 2026 | ISSN 2695-9631
International Visual Culture Review / Revista Internacional de Cultura Visual
https://doi.org/10.62161/revvisual.v18.6137

DEEPFAKES AND *FACT-CHECKING*
Influence of AI on the Disinformation Ecosystem from the Perspective of Fact-checkers and Academics

DAVID GARCÍA-MARÍN [1]
david.garciam@urjc.es
AMAYA NOAIN-SÁNCHEZ [1]
amaya.noain@urjc.es
GUIOMAR SALVAT-MARTINREY [1]
guiomar.salvat@urjc.es
[1] Universidad Rey Juan Carlos, España

KEYWORDS
Artificial Intelligence
Deepfakes
Disinformation
Fact-checking
Fake news
Journalism
Information verification

ABSTRACT
This study aims to analyse the uses and impact of artificial intelligence—particularly deepfakes—in disinformation campaigns, as well as the utility and potential of this technology in combating disinformation. Semi-structured interviews were conducted with 16 fact-checkers and experts in AI and disinformation in the Spanish context, and the data were analysed using grounded theory. The findings highlight how deepfakes play both a symbolic and an operational role in the disinformation ecosystem. Their influence extends beyond the generation of disinformation to affect the broader social perception of reality through strategies such as false-flag marketing and phenomena including the liar's dividend and enshittification.

Received: 10/ 11 / 2025
Accepted: 19/ 01 / 2026

1. Introduction and Theoretical Framework

1.1. Artificial Intelligence and Disinformation Images: Deepfakes

In its current configuration, the phenomenon of disinformation entered public opinion alongside the term *fake news*, a controversial yet rapidly popularised and accepted concept (Wardle, 2019, 2017). It was in 2016, during the Brexit campaign in the United Kingdom and the US presidential elections, that political representatives began using the term to criticise media outlets that were not aligned with them (Bastos & Mercea, 2019; Quandt et al., 2019). This occurred against a backdrop of declining credibility in journalism, where traditional news media had already lost part of their influence and their predominant role as gatekeepers (Allcott & Gentzkow, 2017), while manipulated content on social media—masquerading as journalistic information—proliferated, heightening the climate of confusion (Tandoc et al., 2018). Since then, the disinformation ecosystem has continued to expand, particularly following the Covid-19 pandemic (Salaverría et al., 2020), when a quantitative leap occurred within an increasingly complex context marked by information overload or "infoxication" (WHO, 2020), further propelled in recent years by the popularisation of artificial intelligence (hereinafter, AI).

Beyond quantitative growth, disinformation has also evolved qualitatively. Advances in so-called generative artificial intelligence have enabled the development of algorithms capable of producing false content with the appearance of news (Flores-Vivar, 2020) and realistic audiovisual material. Their implementation on digital platforms to attract users has contributed to the rapid dissemination of hoaxes and deceptive narratives (Bontridder & Poullet, 2021). Among these contents, the most striking use is the creation of deepfakes or hyper-realistic content (Chesney & Citron, 2019), consisting of images—whether photographs or videos (and to a lesser extent audio)—in which individuals (especially prominent figures) appear "performing actions or delivering speeches they never carried out or uttered" (García-Marín, 2025, p. 79). These imitations may have innocuous purposes, such as the BuzzFeed video showing former US President Barack Obama uttering offensive words, which quickly went viral (Vaccari & Chadwick, 2020). Although the initial intent of such videos may not be disinformative, they can induce error or deception when consumed out of context, hindering accurate decoding by audiences.

Indeed, these "algorithmic factories" (García-Marín, 2025) for producing images, even when not intended for disinformation, generate content that can lead to confusion for various reasons. First, both the training data and the design of the algorithms may contain biases that result in discriminatory or offensive content towards certain social groups. Studies such as that of García-Ull and Melero-Lázaro (2023) demonstrate that generative AI systems not only replicate gender stereotypes but reinforce and amplify them. Second, some of the content used by these systems may be protected by copyright, raising legal and ethical risks related to the misappropriation of intellectual property. Finally, these technologies rely on data circumscribed to a specific temporal moment, limiting their capacity to provide updated and relevant content in dynamic contexts.

In any case, the proliferation of deepfakes—whether intended to disinform or not—challenges the composition of a healthy information ecosystem and public sphere. The alteration of audiovisual content through such technologies opens the door to new forms of citizen manipulation (Hancock & Bailenson, 2021; Köbis et al., 2021). Their harmful potential became evident in the early years of their popularisation, when they were massively used to generate fake nude images of women (Hao, 2020) and place them in pornographic videos (Cook, 2019). Subsequently, their use in political contexts—for example, in massive disinformation campaigns targeting voters before elections (Dobber et al., 2021)—represented a qualitative leap (Vaccari & Chadwick, 2020). Undoubtedly, their capacity to further erode trust in public institutions and the media themselves is a cause for concern, as it may undermine the foundations of democracy (Kalpokas & Kalpokiene, 2022; Wahl-Jorgensen & Carlson, 2021).

The primary characteristic of these productions is their hyper-realistic quality, which enhances their credibility among audiences with lower levels of media and information literacy. Recent studies examining the impact of deepfakes—conducted in different contexts and using quasi-experimental methods—show that even when citizens believe they can detect manipulated videos, their ability to distinguish between false and true content is not particularly high (García-Marín et al., 2025; Köbis, et al., 2021). The persuasive power of audiovisual material, far greater than that of text-based hoaxes, combined with its rapid dissemination, lends legitimacy to its content, making disinformation appear more plausible (Tahir et al., 2021). Furthermore, the fact that AI tools for generating deepfakes are available online and easily accessible facilitates their proliferation.

Initial research on deepfake-generated disinformation reveals that the most frequent narratives focus on manipulating the public image of prominent figures, particularly political actors (García-Marín, 2025). Likewise, intensive use of these technologies has been documented in wartime contexts, where fake images are generated to demonise the enemy, justify one's own actions, and boost morale, in line with historical propaganda strategies. Another worrying area is the creation of such content (often linked to foreign campaigns) aimed at destabilising specific collectives, communities, or social groups in certain countries. These AI-fabricated false productions range from photographs showing fake attacks carried out by migrants to videos with fraudulent medical recommendations, with the aim of sowing chaos, fostering hatred, and eroding institutional trust. Economic motivations have also been identified in the generation of this type of disinformation, such as the deceptive promotion of products or financial fraud through false endorsements by influential figures. Finally, the accessibility of deepfake creation tools has facilitated the proliferation of spectacular content with no apparent ideological load, oriented towards maximising virality and economic profit on social media, particularly among professionals in the visual field.

1.2. The Role of Fact-Checkers in Countering Deepfakes

Alongside the increase and growing sophistication of disinformation, platforms dedicated to fact-checking have undergone significant evolution. These entities emerged in the United States in the 2000s to verify the accuracy of statements made by political actors (Graves, 2016) and now examine all types of disinformative content. They detect content ranging from text, images, audio, and videos produced using traditional video-editing techniques (cheapfakes) to more elaborate material generated through algorithmic methods such as deepfakes (Maslej et al., 2023). These platforms may operate as independent organisations or be affiliated with media outlets (Graves et al., 2020). In either case, their work is characterised by collaboration between journalists and various professionals from the technological domain, such as experts in open-source intelligence (OSINT) (Gregory, 2021), software developers, and specialists in digital forensic analysis (Ciampaglia, 2018), among others. This collaboration results in a strong capacity to adapt to technological changes.

For fact-checkers, the integration of AI-driven tools represents a strategic resource for addressing the proliferation of disinformation and, specifically, for detecting deepfakes (Brandtzaeg et al., 2018). From the application of algorithms trained with machine learning and deep learning techniques to automatically identify disinformative content, to chatbots that enable interaction with users who encounter suspicious material (Arias-Jiménez et al., 2023), AI promises to be highly valuable in verification tasks and in preventing the spread of disinformation (Pasquetto et al., 2022). When combined with visual forensic analysis techniques (Gregory, 2021), this work could prove essential—for example—in electoral processes, crisis situations such as that triggered by the Covid-19 pandemic (Luengo & García-Marín, 2020), or geopolitical events such as Russia's invasion of Ukraine (O'Connor, 2022). Nevertheless, some uncertainties remain regarding the actual effectiveness of AI tools in the automatic detection of false content and in the practical implementation of these tools across the different stages of the verification process.

Although preliminary studies conducted in very specific contexts indicate that the impact of deepfakes on fact-checkers' verification routines appear limited and that, for the time being, visual manipulations such as decontextualisations are more difficult to verify (Weikmann et al., 2023), it is foreseeable that the production of disinformation using AI tools will progressively improve. This is precisely why the capacity to adapt to technological changes positions fact-checkers as a key barrier against AI-driven visual disinformative content

2. Objectives and Methodology

In this context, the central objective of this research is to understand the influence of artificial intelligence—particularly deepfakes—on the disinformation ecosystem from the perspective of fact-checking professionals and researchers specialising in disinformation and artificial intelligence. As in previous studies (Gutiérrez-Caneda & Vázquez-Herrero, 2024; Sánchez González et al., 2022), we consider these profiles to be key informants for analysing this object of study, as they are on the front line in the fight against the phenomenon of disinformation.

In this study, the analysis of the influence of these algorithmic systems is broken down into four elements, which will be addressed separately but in an integrated manner: (1) the volume and complexity of AI-generated disinformative content (deepfakes), (2) the uses of AI in disinformation campaigns, (3) the impact of AI on the disinformation ecosystem, and (4) the utility and potential of AI in combating disinformation. Taking these four aspects into account, the specific objectives of the work are focused on:

O1. Analysing the volume and complexity of AI-generated disinformation (deepfakes) from the perspective of verifiers and researchers.

O2. Understanding the uses of AI in disinformation campaigns, according to the views of verifiers and researchers.

O3. Evaluating the impact of AI on the disinformation ecosystem, also from the perspective of the aforementioned experts.

O4. Analysing how verifiers and academics assess the utility of AI tools in the fight against disinformation.

This research, which is eminently qualitative in nature, is based on the principles of grounded theory as proposed by Strauss and Corbin (2002), adopting an inductive approach oriented towards the generation of categories and the emergent understanding of the data. Grounded theory not only seeks to identify the key dimensions and categories of the phenomenon under study, but above all to establish relationships between them. This methodology enables a systematic analysis of the participants' experiences and perceptions without relying on pre-established hypotheses (Salvat-Martinrey et al., 2024), thereby facilitating the identification of patterns in the subjects' discourse (Ahumada et al., 2025).

As a methodological framework, Moscovici's (1979) theory of social representations was employed, which provides the basis for interpreting how participants construct and communicate meanings about a complex phenomenon. This approach allowed exploration of how the interviewed subjects develop knowledge, values, and beliefs regarding the influence of AI on disinformation.

2.1. Instrument and Participants

A total of 16 key informants were interviewed semi-structurally (11 fact-checkers belonging to Spanish verification agencies and 5 researchers specialising in AI and disinformation affiliated with Spanish universities) (Table 1). The inclusion of two distinct profiles (fact-checking professionals and academics) allowed for complementary perspectives on the object of study, although it required adapting the specific interview guide to each profile. It should be noted that the interviews with verification professionals were not limited to journalists or editors but also included members of the engineering and data departments within the verification agencies.

The interviews revolved around the four dimensions outlined in the objectives: (1) the volume and complexity of AI-generated disinformation, (2) the use of AI in disinformation campaigns, (3) the impact of AI on the disinformation ecosystem, and (4) the utility of AI in combating disinformation. Priority was given to conducting the interviews in person, and they were recorded only in audio format. However, for operational reasons, a substantial proportion had to be carried out remotely using Microsoft Teams software, in which case both video and audio were recorded. All interviews were conducted synchronously. They took place between February and June 2025, with durations ranging from 30 to 75 minutes.

Table 1. Profile of the subjects interviewed.

Number of interviews	Profile	Format
ENT_01	University lecturer and researcher in AI and disinformation. With previous professional experience as a *fact-checker.*	Remote
ENT_02	University lecturer and researcher in AI, disinformation and the social impact of algorithms.	Remote
ENT_03	University lecturer and researcher in AI and disinformation. With previous professional experience as a *fact-checker.*	Remote
ENT_04	University lecturer and researcher in AI and disinformation.	Remote
ENT_05	Journalist (*fact-checker*) in Spanish verification agency.	Remote
ENT_06	Journalist (*fact-checker*) in Spanish verification agency.	In-person
ENT_07	Journalist (*fact-checker*) in Spanish verification agency with technological profile.	In-person
ENT_08	Journalist (*fact-checker)* in Spanish verification agency.	In-person
ENT_09	Journalist (*fact-checker*) in Spanish verification agency with technological profile.	In-person
ENT_10	Member of the engineering and data team in Spanish verification agency.	Remote
ENT_11	Member of the engineering and data team in Spanish verification agency.	Remote
ENT_12	Journalist (*fact-checker)* in Spanish verification agency.	Remote
ENT_13	University professor and researcher in AI and disinformation. With previous professional experience as a *fact-checker.*	Remote
ENT_14	Journalist (*fact-checker)* in Spanish verification agency.	In-person
ENT_15	Journalist (*fact-checker)* in Spanish verification agency.	In-person
ENT_16	Journalist (*fact-checker)* in Spanish verification agency.	In-person

Source: Own elaboration, 2025.

2.2. Data Analysis

An inductive approach was employed, utilising the procedures of open and axial coding. The open coding process began with the transcription and reading of the interviews with the aim of deepening understanding of emerging categories and progressing towards theoretical saturation, in line with the constant comparative method characteristic of grounded theory. Units of meaning (words or phrases) with conceptual relevance within the discourse were identified in accordance with the stated objectives. These units were assigned descriptive labels that served as initial markers of the content. Subsequently, these labels were grouped into analytical codes based on observed thematic or conceptual similarities in the data. Each code was supported by verbatim quotations and analysed according to its grounding, that is, the frequency with which it appeared in the discourse, thereby enabling the identification of recurrent patterns in the analysed corpus.

In a second phase, axial coding was carried out. Once the codes had been defined during open coding, they were organised into hierarchical categories using a paradigmatic matrix (Strauss & Corbin, 2002). This procedure allowed the identification of causal, conditional, and contextual relationships between the codes. The established relationships included links such as "is part of", "explains", "is characterised by", or "is a", thereby facilitating conceptual integration. Given the visual complexity of the code networks, the results section of this article presents summarised and simplified versions of these networks. The complete original versions may be consulted at: https://figshare.com/s/a7b4218c75be1a905ee4.

Finally, the categories and subcategories were systematically reviewed with the objective of identifying a theoretical thread that would account for the social representation of each of the dimensions analysed, which are presented in the following section.

For the presentation of results, participants were anonymised. Accordingly, excerpts extracted from the interviews are presented using the following formula: ENT_XX, where XX is the interview number assigned in Table 1.

Data processing was conducted using the specialised qualitative research software Atlas.ti v.25. The study received approval from the Ethics Committee of the researchers' university (internal registration number: 120120250322025).

3. Results

3.1. Volume and Complexity of AI-Generated Misinformation

AI-generated disinformation, although increasingly sophisticated, does not yet appear to be predominant. In a sense, study participants downplay the current impact of this technology on disinformation production, at least in quantitative terms. They note that "we are not yet in the realm of large-scale images or major hoaxes because the technology has not become very sophisticated yet" (ENT_13). Moreover, they recognise that "for the disinformer, it is still more cost-effective to spread an out-of-context video than to create one with AI" (ENT_12), suggesting that, for now, traditional strategies and technologies remain more profitable and effective. As a result, the volume of AI-generated disinformative content that verifiers must address is not particularly high, at least "not in as massive a way as we initially expected, since at first there were more alarmist voices" (ENT_07).

In this regard, the interviewed experts agree that AI-generated falsehoods do not yet constitute the main source of disinformation. One of the experts interviewed points out that "according to data from the European Digital Media Observatory (EDMO), around 5 to 10% of verifications are based on AI-generated hoaxes" (ENT_12), although he acknowledges that "this percentage was equal to or less than 1% just a few years ago" (ENT_12), indicating a growing trend. However, it is qualified that "we are not seeing a large wave of AI-generated falsehoods nor a great deal of elaboration, because cheapfakes already work, and that is why there is no need to elaborate them further" (ENT_01). Cheapfakes—crude manipulations (Gamir-Ríos & Tarullo, 2022) that combine simple techniques such as inadequate contextualisation or retouching with image-editing tools (Schick, 2020)—have become an effective and accessible strategy, leaving little room for more sophisticated and complex disinformative content: "There are indeed many cheapfakes, many videos where anyone slightly manipulates lip movement, adds a cloned synthetic voice […] and we have seen quite an increase in that" (ENT_07). The accessibility of tools capable of generating this type of content has democratised their use for disinformative purposes, although it has also limited, for the time being, their technical complexity in general terms.

In parallel, it seems evident that AI-generated disinformation is in a phase of expansion and diversification, characterised by progressive improvement in technical quality, growing accessibility, and partial integration into the dynamics of information manipulation: "We see it being used more and more, that access is very easy, that it is becoming popular, that there is a series of trends in which usage is very much on the rise" (ENT_06). Although it does not yet dominate the disinformation ecosystem, its disruptive potential generates some concern. As one expert concludes: "I think it is going to get complicated for all of us, […] Google's Veo3 is already impressive in what it does" (ENT_14).

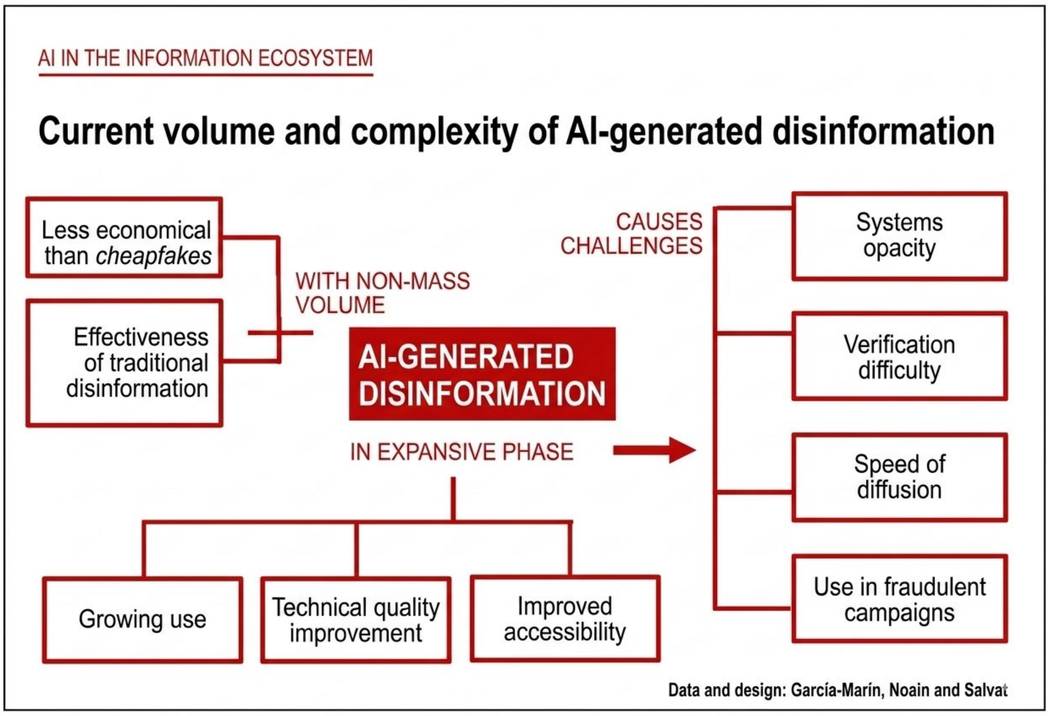
Along these lines, analysis of the interviews with verifiers and experts confirms that the emergence of algorithmic tools for creating disinformation has introduced new complexities into the contemporary information ecosystem. The increasing sophistication of AI-generated disinformation poses not only technical challenges but also epistemological and methodological ones for fact-checking. One of the identified challenges is the opacity of AI systems, described as "black boxes" (ENT_01), an aspect that hinders the traceability of content. As one of the study participants notes, "it is very difficult to verify where that disinformation comes from, because you cannot trace the origin as you could in another context" (ENT_01). In the same vein, another relevant aspect is the methodological difficulty involved in verifying content generated entirely with AI. The experts explain that "the fact that it is created from scratch with an AI tool makes it difficult to trace the origin in many cases" (ENT_08). This lack of transparency compromises one of the fundamental pillars of verification: the identification of the original source. Unlike traditional manipulations, which started from a real image or video, content generated entirely by AI lacks a verifiable referent, which requires rethinking analysis strategies.

In terms of quality, a significant evolution is also observed. Although low-quality content persists, "there are others that are increasingly of higher quality and are becoming more similar to real content" (ENT_03). This technical improvement increases the risk of identity impersonation, particularly in the case of public figures. One of the experts warns: "What scares me the most are those contents that are going to impersonate the image of a relevant person and put words in their mouth that they never said" (ENT_05). This type of audiovisual manipulation, by exploiting the credibility of moving images, makes subsequent refutation difficult, even when evidence to the contrary is available.

In summary, it is not only a question of how much disinformation is generated with AI, but of how it is configured and what kind of challenges its detection and debunking present. The interviewed subjects conclude that "what makes it difficult for us is the typology, more than the quantity" (ENT_07). This complexity is aggravated by the speed of dissemination and the ease of access to generative tools, which has favoured their use in fraudulent campaigns, such as "the promotion of fake cryptocurrencies, investment platforms and false medicines" (ENT_06).

Figure 1 shows the network of codes on the volume and complexity of disinformation produced with AI.

Figure 1. Code network regarding the volume and complexity of AI-produced disinformation



Source: Own elaboration, 2025.

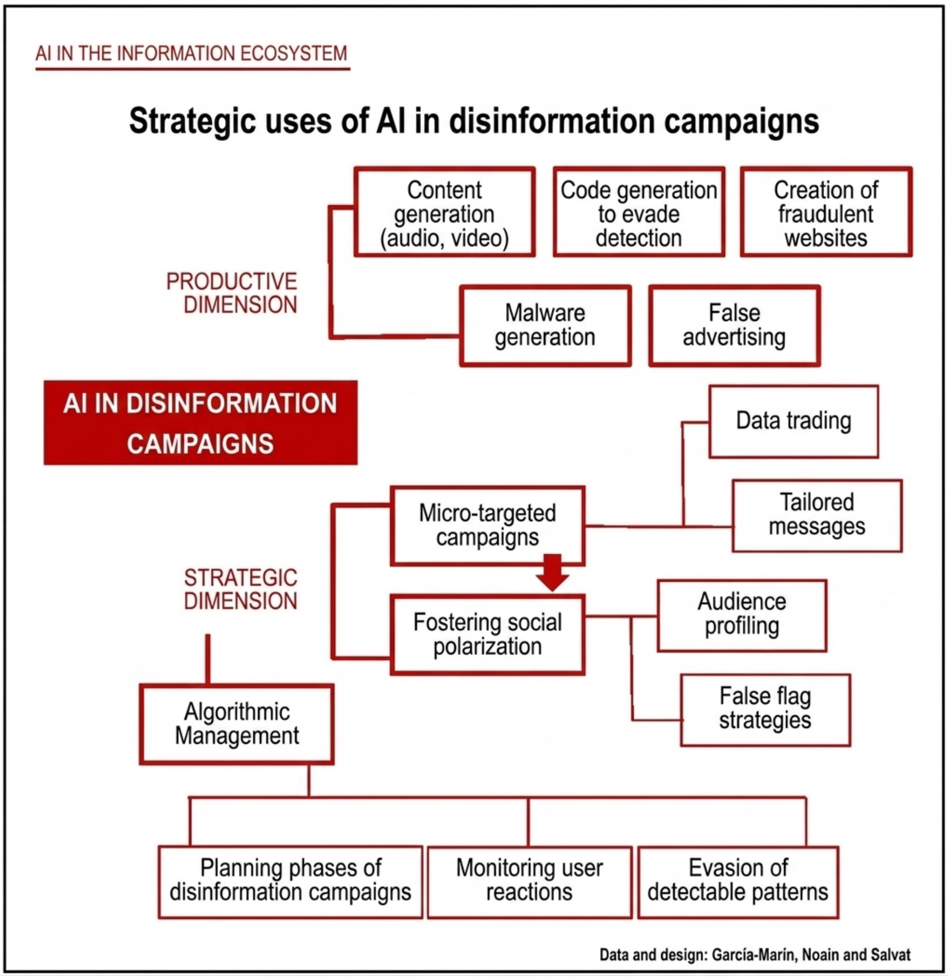
3.2. Uses of AI in Disinformation Campaigns

Regarding the use of AI as a disinformative tool (increasing, as explained in the previous section), two clearly differentiated dimensions are perceived: (1) productive and (2) strategic (Figure 2). The productive dimension refers to cases where AI directly generates some type of content intended to convey a disinformative message or to serve as support for such messages. It therefore stands out not only for the direct generation of images and audio—the production of text with AI is considered residual—but also for the use of these systems to automatically generate code that prevents AI-produced false content from being detected, such as the creation of systems to block the archiving of fraudulent pages. Likewise, the use of AI has been documented in phishing campaigns, malware generation, the creation of fraudulent websites, and fake advertisements. These applications show that AI affects not only the symbolic level but also the operational one, facilitating large-scale fraud and manipulation.

The strategic uses of AI in disinformation campaigns are less visible and more complex to detect. First, the use of these tools for creating highly personalised disinformation campaigns has been documented. One of the experts explains that "companies operating on the dark web sell data on individuals with linked personality and emotional profiles, which allows messages to be generated tailored to each person" (ENT_04). These campaigns are algorithmically managed with AI tools, including the planning of their different phases, such as monitoring user reactions and evading detectable patterns that would facilitate identification: "When to release disinformative messages is something done with algorithmic support, so that verifiers and researchers do not find an established pattern" (ENT_04).

AI is also used to foster social polarisation, for example through false-flag strategies. The so-called false-flag marketing consists of a communicative tactic in which an entity—whether an organisation, group, or individual—presents itself as another, usually as a neutral party or even as an antagonist, with the purpose of shaping public opinion, inducing certain behaviours, or discrediting an adversary. This concept originates in the military and intelligence fields, where it refers to covert operations designed to appear as actions carried out by the enemy. In this type of operation, disinformers use AI "to produce controversial content on social networks in order to create polarity and to detect users at both poles" (ENT_04), which enables profiling of individuals and directing disinformation campaigns adapted to each profile, also managed with AI, thereby amplifying the fragmentation of the public sphere.

Figure 2. Code network on the uses of AI in disinformation campaigns.



Source: Own elaboration, 2025.

3.3. Impact of AI on the Disinformation Ecosystem

According to the participants in the study, AI rarely serves as the central element of disinformation campaigns; rather, it is used to reinforce existing narratives, generating a bandwagon or visibility effect for campaigns launched through other types of productions (cheapfakes) (Figure 3). In this role of supporting pre-existing disinformative narratives, AI-generated content lends them "more realism, greater reach" (ENT_12), and enables "sowing doubt or creating conversational frames around certain topics that are of interest" (ENT_08).

One of the most concerning aspects highlighted by the experts is AI's capacity to produce highly persuasive content, even when false. As one warns, "generative artificial intelligence has nothing to do with truth, but with rhetoric and persuasion, since it is highly convincing even when it is wrong" (ENT_01). This characteristic, combined with the phenomenon of hallucinations—fabricated content without factual basis—has led some researchers to refer to AI as artificial ignorance, that is, a form of automated ignorance that produces errors with an appearance of veracity.

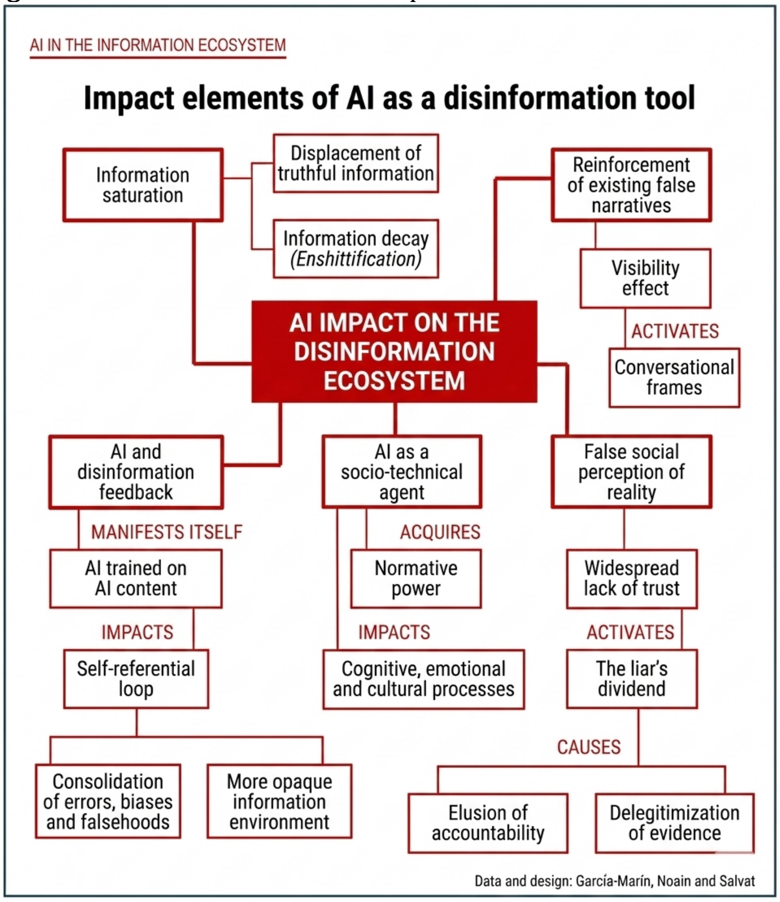
The impact of this technology is not limited to the generation of disinformation; it also affects the broader social perception of reality. Uncritical use of AI can "worsen citizens' trust in information" (ENT_01) and foster generalised distrust that leads to the rejection of even truthful content. For the study participants, the cumulative effect of these phenomena results in a growing sense of unreality: "People are becoming more sceptical, [...] we even doubt real news now" (ENT_13). This situation gives rise to phenomena such as the liar's dividend, a communicative paradox that emerges when the proliferation of fake news and systematic disinformation campaigns erode public trust in the veracity of any content, even the authentic. This phenomenon confers a strategic advantage on those who lie deliberately, since, amid generalised suspicion about the authenticity of information, any individual can deny true facts by claiming they are forgeries or manipulations. This allows dishonest actors to exploit the scepticism generated by disinformation to evade accountability, delegitimise real evidence, or sow doubt about verifiable facts.

Along the same lines, one of the most frequently mentioned risks is AI's capacity to flood the information space with false content, displacing truthful information. "It can be created so quickly that a point may be reached where, when we search for something on platforms like the internet, more disinformative content made with AI appears than real and informative content" (ENT_03). This phenomenon, known as enshittification, involves a progressive degradation of the digital ecosystem, where the overabundance of AI-generated content can make it difficult to distinguish between what is true and what is false, or between high-quality content and that which lacks it.

On the other hand, the experts point out that AI acts not only as a technological tool but also as a socio-technical agent capable of influencing the cognitive, emotional, and cultural processes of contemporary societies. From this more structural perspective, the experts warn that AI systems must be understood as cognitive agents that do not merely process information, but "influence the way we construct knowledge, and the way we relate experiences to emotions" (ENT_02). In this sense, AI acquires a normative power that transcends its instrumental function, sometimes acting as "a police officer, a judge, a journalist, or a doctor" (ENT_04), that is, as a symbolic actor with the capacity to shape and transform social reality.

Ultimately, an emerging phenomenon is highlighted: the feedback loop between AI and disinformation. Until now, "AI was being trained with content produced by humans, but from now on it will take in many documents generated by AI itself" (ENT_15). This self-referential loop can consolidate errors, biases, and falsehoods, creating an increasingly opaque and self-reproducing information environment.

Figure 3. Network of codes on the impact of AI as a disinformative tool.



Source: Own elaboration, 2025.

3.4. Utility of AI in Combating Disinformation

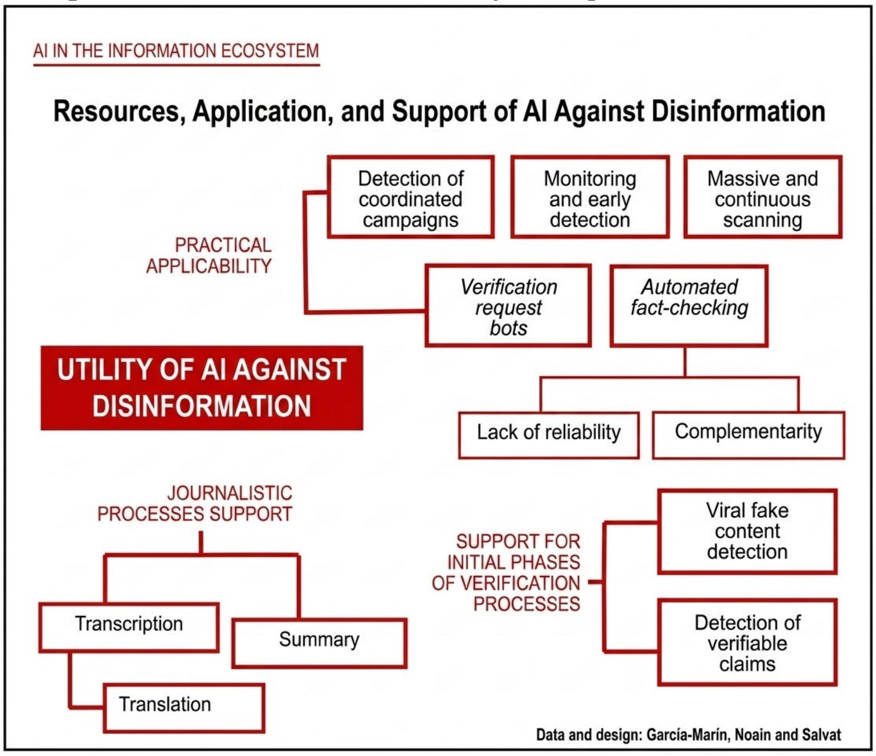
Alongside the negative impact that AI can exert on the current information ecosystem, the potential offered by this technology to combat disinformation campaigns is evident. These systems represent a valuable but limited tool within the verification ecosystem, particularly regarding the automated checking of potentially false content. Their utility lies in the automation of repetitive tasks (or those requiring the automatic processing of large volumes of data), the expansion of monitoring scope, and the identification of emerging patterns. However, their effectiveness is conditioned by technical, epistemological, and structural factors that, as discussed below, prevent full autonomy.

The interviewed subjects agree that these algorithmic models are efficient in the initial phases of the verification process, such as the detection of viral false content or potentially verifiable claims, as well as providing support in processes already present in other journalistic modalities, such as summarisation, translation, or transcription of texts (Figure 4). However, they express greater doubts about their effectiveness in the content verification phase. It can therefore be inferred that, to date, the introduction of so-called automated fact-checking remains in a phase not free of tensions, limitations, and structural challenges that condition its implementation. It should be borne in mind that this type of verification consists of the application of computational tools to assess the veracity of potentially false content through algorithms that analyse it in depth, cross-reference it against databases, documentary corpora, or verified sources, and issue a judgement on its degree of truthfulness, all without direct human intervention (García-Marín, 2022). As will be explained later, the study participants qualify the real scope of these technologies for automated verification.

As mentioned earlier, in terms of practical applicability, AI tools have proven useful in identifying diffusion patterns, analysing narratives, and detecting coordinated campaigns. The instrumental value of AI is recognised in the preliminary phases of the verification process, especially in the monitoring and early detection of potentially disinformative content circulating on digital platforms. Verifiers note that "in the first phase of collecting those data or those disinformation or more viral deceptive contents, for example, AI is starting to help us" (ENT_03). This capacity for massive and continuous scanning enables the expansion of information surveillance, automating tasks that previously required intensive human resources. In this sense, it is highlighted that "AI does not get tired, so the monitoring part, social listening, can be greatly expanded" (ENT_11).

Likewise, some verification agencies have developed specific tools to facilitate the identification of verifiable claims in audiovisual content. An example of this is the use of video software that "automatically offers the verifiable phrases from that audiovisual content" (ENT_11), which optimises the work of verifiers by reducing the time required for preliminary analysis. Systems for narrative analysis have also been implemented that enable the detection of "which narratives are the most powerful, how they evolve, when they rise and when they fall" (ENT_07), proving useful for anticipating structured disinformation campaigns. Similarly, the use of "automatic bots to receive verification requests for hoaxes from the audience" (ENT_08) is noteworthy, a tool that also allows automatic responses to such requests when the requested verification is already available in the fact-checker's database.

Figure 4. Network of codes on the utility of AI against disinformation.



Source: Own elaboration, 2025.

As observed earlier, both verifiers and experts agree on the limited effectiveness of automatic falsehood detection tools. Although applications exist that provide probabilities of content having been artificially generated, "there is no magic tool, as of today, that can tell you whether something is true or false, or generated with AI, with 100% certainty" (ENT_12). These tools also produce inconsistent results, since "sometimes they get it right and sometimes they don't" (ENT_08), which prevents their use as conclusive evidence in verification work. For this reason, it is emphasised that "our research conclusion can never be based solely on what the AI says" (ENT_06). Moreover, their effectiveness depends on constant updating, since "every day 20 new tools come out and you cannot build a toolbox that will still be useful, not even for six months" (ENT_16).

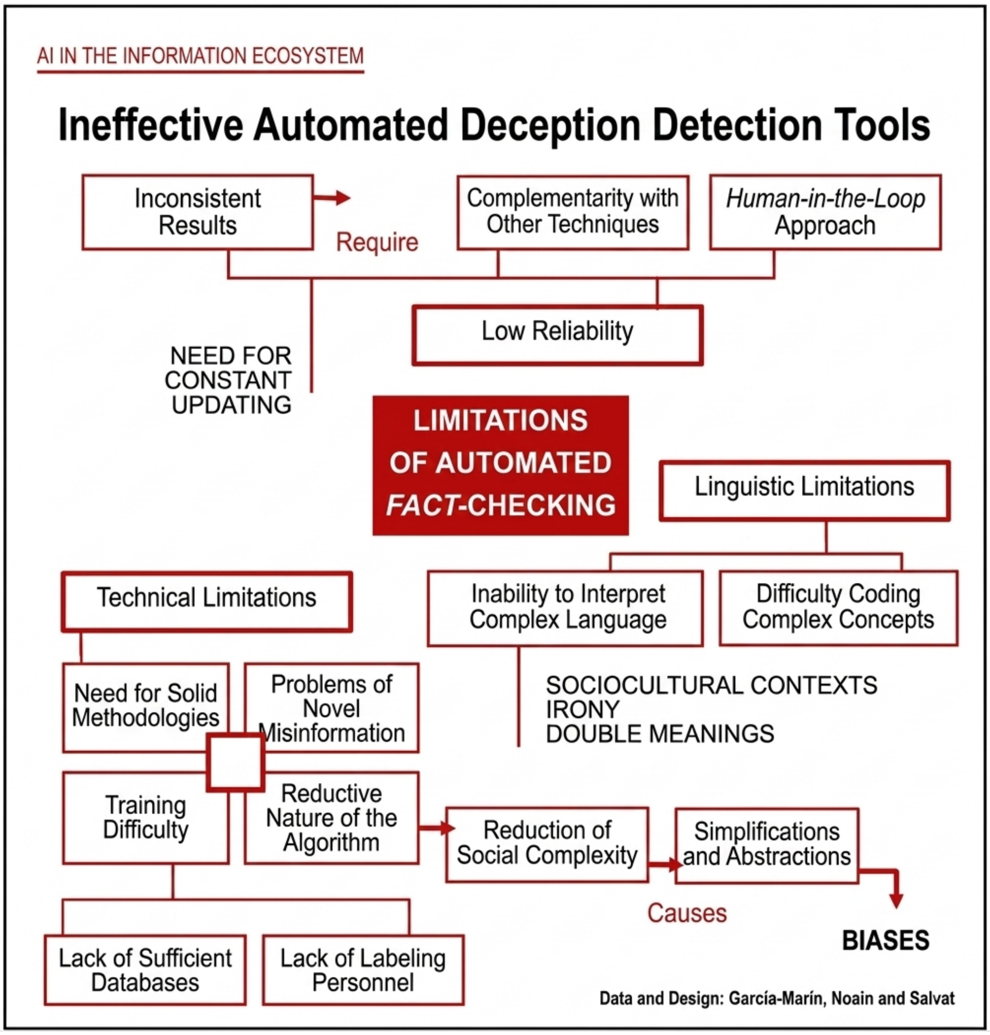
This lack of reliability requires that the predominant approach be human-in-the-loop, where AI acts as support but never as a substitute for journalistic analysis: "The first step is always the analysis by the editor or editors who conducted the investigation, and the AI tool is support, but it is never the basis" (ENT_06). This complementarity is also justified by the inability of these technologies to interpret complex elements of language such as irony, double meanings, or sociocultural contexts. They also struggle to detect novel disinformative content that cannot be compared with previously stored material in verifiers' databases. For all these reasons, the experts agree that AI cannot replace human judgement in the critical phase of verification. As one states categorically, "for verification, if anyone imagines there is a tool that will automatically say whether something is false and no human will be needed to confirm it, they are mistaken" (ENT_01). In the same vein, another interviewee concludes: "The tool may help you arrive sooner or better, but human knowledge and experience are always fundamental" (ENT_09).

This merely complementary role, always in combination with other verification strategies, is determined by a series of limitations that these tools exhibit (Figure 5). First, they present technical difficulties arising from their training. "Sometimes there are not enough databases to train them, and there is not even enough personnel" (ENT_01), which compromises their reliability and thematic coverage. Experts and fact-checkers stress the need to establish solid methodological criteria for their training, as well as adequate data infrastructure for the development of reliable models. "In verification there is no large corpus to train these models, neither in English nor in Spanish" (ENT_13), which limits the development of robust systems. Ideally, multimodal models would be developed that integrate text, image, and context, but this requires "a great deal of human labelling so that the training is correct" (ENT_13).

In addition, structural problems related to access to data and control of the algorithms of the social media platforms where most of the disinformative content checked by verifiers circulates are identified. "We do not have that power as long as we do not have access to the algorithms that are in the hands of a few tech corporations" (ENT_02), which generates an asymmetric dependence on the major platforms. This situation is exacerbated by restrictions on API use, as in the case of X (formerly Twitter): "The last time I tried to use the X API to conduct an analysis of hate speech, I was unable to do so" (ENT_03).

Finally, another relevant obstacle is the lack of conceptual consensus on key phenomena closely related to disinformation, which hinders their algorithmic coding. One of the interviewed experts gives the example of hate speech: "If we are not capable of agreeing on what hate speech is, for example, it will be difficult to program a tool to detect this type of message" (ENT_02). This difficulty is amplified by the reductive nature of algorithms, which "minimise social complexity and make simplifications and abstractions" (ENT_11), potentially introducing ideological biases and inaccuracies into the results.

Figure 5. Network of codes on the limitations of automated fact-checking.



Source: Own elaboration, 2025.

4. Discussion and Conclusions

The research has made it possible to analyse the volume and complexity of AI-generated disinformation (O1), revealing that although this type of content is not yet predominant, its ease of use, technical evolution, and growing accessibility anticipate a significant increase in its utilisation. This phenomenon becomes particularly concerning in contexts of high polarisation or strategic interest, where deepfakes could assume a leading role. Despite initial assumptions that foresaw a massive wave of synthetic content for disinformation (Chesney et al., 2019), the verifiers, experts, and academics interviewed in this study downplay its current frequency, in line with what has been described by Kalpokas and Kalpokiene (2022). At least in quantitative terms, the impact of this technology remains limited, although an upward trend is discernible. This low proliferation is grounded in two aspects: (1) there is a lag between the technical potential of AI and its practical implementation in disinformation campaigns, and (2) cheapfakes are more accessible and cost-effective, which is why they continue to be the most frequent manipulations (Gamir-Ríos & Tarullo, 2022).

Nevertheless, the fact that its use is not currently concerning in quantitative terms does not mean its future disinformative potential should be underestimated. The shared perception among the interviewees suggests that the problem lies not so much in the quantity of deepfakes as in the qualitative leap they represent and the methodological challenges they pose for verification. The fact that generative AI is transforming the production of false content, making it more plausible and harder to detect (Bontridder & Poullet, 2021; Flores-Vivar, 2020), anticipates a more complex landscape for the work of verifiers in a context of low credibility where audiences struggle to identify AI-elaborated false content (García-Marín et al., 2025; Köbis et al., 2021).

Regarding the uses of AI in disinformation campaigns (O2), the findings show that deepfakes primarily operate as narrative reinforcement elements rather than as the core of campaigns. This amplifying function increases their capacity to sow doubt and legitimise false discourses, especially when integrated into algorithmic dissemination strategies. Likewise, AI is employed both in the direct generation of content and in the strategic planning of its circulation, reinforcing its operational and symbolic role in the disinformation ecosystem.

The study's findings reinforce the warnings of Bontridder and Poullet (2021) regarding AI's role in the personalisation of disinformation campaigns and in the algorithmic planning of their dissemination. The operational dimension of AI, evidenced in its use for generating malware, fraudulent pages, or phishing campaigns, is complemented by a more sophisticated and less visible dimension: the algorithmic management of highly personalised campaigns. Companies operating on the dark web commercialise psychological and emotional profiles that enable disinformative messages to be tailored to each individual, optimising their persuasive impact. This planning includes monitoring reactions and evading detectable patterns, which complicates the work of verifiers and challenges traditional analysis methods.

In addition, AI is used as a tool to foster social polarisation (Torcal & Magalhaes, 2022) through strategies such as false-flag marketing (García-Marín, 2025). These techniques allow disinformers to simulate antagonistic identities to shape public opinion and discredit adversaries. AI thus facilitates the production of controversial content on social networks with the aim of detecting users at both ideological poles, profiling their characteristics, and directing adapted disinformation campaigns to each group.

Our work also underscores that the impact of AI (O3) extends beyond the mere production of disinformative content. One of the most relevant challenges identified in this study is the opacity of artificial intelligence systems, described by participants as genuine "black boxes". This characteristic significantly hinders the traceability of false content (Morosoli et al., 2025; Wu, 2024), as it prevents tracing its origin with the precision possible in other contexts. This technical and epistemological limitation compromises one of the fundamental pillars of fact-checking: the identification of primary sources and the reconstruction of the content production process (Graves, 2016).

The lack of transparency in algorithmic systems is linked to a broader concern about the erosion of truthfulness in the digital environment, information infoxication, and so-called enshittification, which erode public trust and make it difficult to distinguish between what is true and what is false. In this sense, our study aligns with the reflections of Kalpokas and Kalpokiene (2022) on epistemic anarchy, in which the criteria for determining what is true become diluted, generating a form of artificial ignorance that describes how AI can produce erroneous content with an appearance of veracity. AI directly affects the social perception of reality, fostering generalised distrust that extends not only to false content but also to truthful content. This situation reinforces the phenomenon of the decay of truth described by Chesney and Citron (2019) and contributes to the so-called liar's dividend, in which generalised distrust enables dishonest actors to deny even true facts. In sum, the feedback loop between AI and disinformation—that is, the possibility that algorithmic systems will be trained with content generated by AI itself—poses risks inherent in the consolidation of errors and biases (Deuze & Beckett, 2022), which could further aggravate the opacity of the information environment and make verification work even more difficult.

Finally, regarding the fourth objective—analysing how verifiers and academics assess the utility of AI tools in combating disinformation (O4)—it is concluded that these tools have a valuable but limited role. This perception confirms the results of previous research. For some authors, the advent of AI has increased the possibilities of combating disinformation (Moreno Espinosa et al., 2024; Rubin, 2022) and can help distinguish between truthful information and the distortion of reality (Flores Vivar, 2019; Santos, 2023) or even reduce detection time and increase responsiveness to disinformation campaigns (Cuartielles et al., 2024).

This potential of AI proves especially useful in the initial phases, that is, in tasks of monitoring false content and detecting disinformation campaigns, as well as in the automation of processes. Its application to improve the productivity of fact-checkers is therefore highlighted (Cools & Diakopoulos, 2024; Cools & de Vreese, 2025; Dodds et al., 2025; García de Torres et al., 2025; Wu et al., 2018). Nevertheless, it is emphasised that these tools cannot replace human judgement, owing to their limited reliability, their inability to interpret complex contexts, and the lack of adequate data for their training. This instrumental or complementary view of AI reinforces the human-in-the-loop approach (García-Marín, 2022), where AI acts as support but not as a substitute for journalistic analysis. Thus, although its utility in monitoring and preliminary analysis tasks is recognised, the interviewees agree that automated fact-checking still presents technical and epistemological limitations. This view aligns with what has been proposed by García-Marín (2022), who warns that automated verification cannot replace human judgement, particularly in complex contexts.

Likewise, the effective implementation of AI in fact-checking requires overcoming technical, epistemological, and structural obstacles, such as the scarcity of adequate corpora, restricted access to data and algorithms of digital platforms, and the lack of conceptual consensus on key phenomena such as hate speech.

4.1. Limitations of the Study

This work presents three methodological limitations. First, the sample is confined to the Spanish context, which limits the extrapolation of results to other sociocultural and media environments. Although 16 key informants were interviewed (11 fact-checkers and 5 researchers), all belong to Spanish agencies and universities. This may imply a geographical and cultural limitation that would affect the generalisation of the results to other international contexts, where the use of AI and disinformation dynamics may differ significantly.

Second, the research is based on qualitative interviews, which allow for a rich exploration of meanings but also entail a certain dependence on individual perceptions. As is common in this type of work, the opinions of the participants could be influenced by their personal experiences, professional biases, or the specific temporal moment at which the interviews were conducted.

Finally, the methodological design relies exclusively on semi-structured interviews, without incorporating other qualitative techniques (such as focus groups, observational methods, or documentary analysis) or quantitative methods that would enable the findings to be contrasted or complemented. Future research, based on these methodological approaches, could complement the interpretive depth of the results obtained in our study.

5. Acknowledgements

This research is funded by the project "Challenges, uses and limitations of AI in fact-checking and the fight against disinformation" (DESAF_IA) (Ref. 2024/SOLCON-135623) funded by the call for Projects IMPULSO to research of the Rey Juan Carlos University (2024).

References

Asiri, A., Panday-Shukla, P., Rajeh, H. S., & Yu, Y. (2021). Broadening perspectives on CALL teacher education: From technocentrism to integration. *TESL-EJ, 24*(4).

Ahumada, R., Aravena-Winkler, M., & Sacomori, C. (2025). Social representations of interdisciplinary work from the perspectives of students and clinical tutors on a cardiovascular rehabilitation program in Chile. *International Journal of Educational Research Open*, *9*, 100515. https://doi.org/10.1016/j.ijedro.2025.100515

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives, 31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

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, Á. Torres-Toukoumidis, A. De-Santis, & Ó. Avilés (Eds.), *Communication and applied technologies* (Vol. 318). Springer. https://doi.org/10.1007/978-981-19-6347-6_12

Bastos, M. T., & Mercea, D. (2019). The Brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review, 37*(1), 38–54. https://doi.org/10.1177/0894439317734157

Bontridder, N., & Poullet, Y. (2021). The role of artificial intelligence in disinformation. *Data & Policy, 3*(3), 1–21. https://doi.org/10.1017/dap.2021

Brandtzaeg, P. B., Følstad, A., & Chaparro-Domínguez, M.-Á. (2018). How journalists and social media users perceive online fact-checking and verification services. *Journalism Practice, 12*(9), 1109–1129. https://doi.org/10.1080/17512786.2017.1363657

Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review, 107*, 1753–1820. https://scholarship.law.bu.edu/faculty_scholarship/640

Ciampaglia, G. L. (2018). Fighting fake news: A role for computational social science in the fight against digital misinformation. *Journal of Computational Social Science, 1*(1), 147–153. https://doi.org/10.1007/s42001-017-0005-6

Cook, J. (2019, June 23). Here´s what it´s like to see yourself in a deepfake porn video. *The Huffington Post.* https://www. huffingtonpost.co.uk/entry/deepfake-porn-video_uk_5d106e03e4b0aa375f4f1ea7

Cools, H., & de Vreese, C. H. (2025). From Automation to Transformation with AI-Tools: Exploring the Professional Norms and the Perceptions of Responsible AI in a News Organization. *Digital Journalism*, 1–20. https://doi.org/10.1080/21670811.2025.2505982

Cools, H., & Diakopoulos, N. (2024). Uses of Generative AI in the Newsroom: Mapping Journalists' Perceptions of Perils and Possibilities. *Journalism Practice*, 1–19. https://doi.org/10.1080/17512786.2024.2394558

Cuartielles, R., Mauri-Ríos, M. & Rodríguez-Martínez, R. (2024). Transparencia en el uso de la IA en las plataformas de fact-checking en España y sus desafíos éticos. *Communication & Society, 37*(4), 257-271. https://doi.org/10.15581/003.37.4.257-271

Deuze, M., & Beckett, C. (2022). Imagination, Algorithms and News: Developing AI Literacy for Journalism. *Digital Journalism*, *10*(10), 1913–1918. https://doi.org/10.1080/21670811.2022.2119152

Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2021). Do (microtargeted) deepfakes have real effects on political attitudes? *The International Journal of Press/Politics, 26*(1), 69–91. https://doi.org/10.1177/19401612209443

Dodds, T., Zamith, R., & Lewis, S. C. (2025). The AI turn in journalism: Disruption, adaptation, and democratic futures. *Journalism*. https://doi.org/10.1177/14648849251343518

Flores Vivar, J. M. (2019). Inteligencia artificial y periodismo: diluyendo el impacto de la desinformación y las noticias falsas a través de los bots. *Doxa Comunicación. Revista Interdisciplinar de Estudios de Comunicación y Ciencias Sociales*, *29*, 197-212. https://doi.org/10.31921/doxacom.n29a10

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

Gamir-Ríos, J. & Tarullo, R. (2022). Predominio de las *cheapfakes* en redes sociales. Complejidad técnica y funciones textuales de la desinformación desmentida en Argentina durante 2020. *AdComunica*, (23), 97-118. https://doi.org/10.6035/adcomunica.6299

García de Torres, E., Ramos, G., Yezers' ka, L., Gonzales, M., Higuera, L., & Herrera, C. (2025). The use and ethical implications of artificial intelligence, collaboration, and participation in local Ibero-American newsrooms. *Frontiers in Communication*, *10*, 1539844. https://doi.org/10.3389/fcomm.2025.1539844

García-Marín D. (2022). Modelos algorítmicos y fact-checking automatizado. Revisión sistemática de la literatura. *Documentación de las Ciencias de la Información*, *45*(1), 7-16. https://doi.org/10.5209/dcin.77472

García-Marín, D. (2025). *Desinformación y fact-checking en la era de la IA*. Ediciones CIESPAL.

García-Marín, D., Suárez-Álvarez, R., & García-Jiménez, A. (2025). "Todo parece veraz". Credibilidad de la desinformación producida usando IA desde la perspectiva de los estudiantes de comunicación en España. *Revista De Comunicación*, *24*(2), 183–227. https://doi.org/10.26441/RC24.2-2025-3872

García-Ull, F.-J., & Melero-Lázaro, M. (2023). Gender stereotypes in AI-generated images. *Profesional de la información, 32*(5). https://doi.org/10.3145/epi.2023.sep.05

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

Graves, L., & Lauer, L. (2020). From movement to institution: The "Global Fact" Summit as a field-configuring event. *Sociologica, 14*(2), 157–174. https://doi.org/10.6092/issn.1971-8853/11154

Gregory, S. (2021). Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism. *Journalism, 23*(3), 708–729. https://doi.org/10.1177/14648849211060644

Gutiérrez-Caneda, B., & Vázquez-Herrero, J. (2024). Redrawing the Lines Against Disinformation: How AI Is Shaping the Present and Future of Fact-checking. *Tripodos*, (55), 55–74. https://doi.org/10.51698/tripodos.2024.55.04

Hancock, J. T., & Bailenson, J. N. (2021). The social impact of deepfakes. *Cyberpsychology, Behavior and Social Networking, 24*(3), 149–152. https://doi.org/10.1089/cyber.2021.29208.jth

Hao, K. (2020, October 20). A deepfake bot is being used to ''undress'' underage girls. *MIT Technology Review*. https://www.technologyreview.com/2020/10/20/1010789/aideepfake-bot-undresses-women-and-underagegirls/

Kalpokas, I., & Kalpokiene, J. (2022). On alarmism: Between infodemic and epistemic anarchy. In I. Kalpokas & J. Kalpokiene (Eds.), *Deepfakes: A realistic assessment of potentials, risks, and policy regulation* (pp. 41–53). Springer.

Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *iScience, 24*(11), 103364. https://doi.org/10.1016/j.isci.2021.103364

Luengo, M., & García-Marín, D. (2020). The performance of truth: Politicians, fact-checking journalism, and the struggle to tackle COVID-19 misinformation. *American Journal of Cultural Sociology, 8*(3), 405–427. https://doi.org/10.1057/s41290-020-00115-w

Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J. C., Parli, V., Shoham, Y., Wald, R., Clark, J., & Perrault, R. (2023). *The AI Index 2023 Annual Report*. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. https://aiindex.stanford.edu/report

Moreno Espinosa, P., Abdulsalam Alsarayreh, R. A., & Figuereo Benítez, J. C. (2024). El Big Data y la inteligencia artificial como soluciones a la desinformación. *Doxa Comunicación. Revista Interdisciplinar de Estudios de Comunicación y Ciencias Sociales*, *38*, 437-451. https://doi.org/10.31921/doxacom.n38a2029

Morosoli, S., Resendez, V., Naudts, L., Helberger, N., & de Vreese, C. (2025). "I Resist". A Study of Individual Attitudes Towards Generative AI in Journalism and Acts of Resistance, Risk Perceptions, Trust and Credibility. *Digital Journalism*, 1–20. https://doi.org/10.1080/21670811.2024.2435579

Moscovici, S. (1979). *El psicoanálisis, su imagen y su público*. Huemul.

O'Connor, C. (2022). TikTok is at the forefront of Russia's propaganda war. https://www.isdglobal.org/isd-in-the-news/ciaran-oconnor-tiktok-is-at-the-forefront-of-russias-propaganda-war/

Pasquetto, I. V., Jahani, E., Atreja, S., & Baum, M. (2022). Social debunking of misinformation on WhatsApp: The case for strong and in-group ties. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW1), 1–35. https://doi.org/10.1145/3512964

Quandt, T., Frischlich, L., Boberg, S., & Schatto-Eckrodt, T. (2019). Fake news. *The International Encyclopedia of Journalism Studies*, 1–6. https://doi.org/10.1002/9781118841570.iejs0128

Rubin, V. L. (2022). *Misinformation and disinformation: Detecting fakes with the eye and AI*. Springer.

Salaverría, R., Buslón, N., López-Pan, F., León, B., López-Goñi, I., & Erviti, M.-C. (2020). Desinformación en tiempos de pandemia: Tipología de los bulos sobre la Covid-19. *Profesional de la Información, 29*(3), e290315. https://doi.org/10.3145/epi.2020.may.15

Salvat-Martinrey, G., García-Marín, D., & Zorogastua, J. (2024). La inteligencia artificial generativa y su impacto en el sector de la comunicación. Percepción de los futuros profesionales. *RAE-IC, Revista de la Asociación Española de Investigación de la Comunicación*, *11*, raeic11e06. https://doi.org/10.24137/raeic.11.e.6

Sánchez-González, M., Sánchez-Gonzales, H.M., & Martínez-Gonzalo, S. (2022). Inteligencia artificial en verificadores hispanos de la red IFCN: proyectos innovadores y percepción de expertos y profesionales. *Estudios sobre el Mensaje Periodístico, 28*(4), 867-879. https://dx.doi.org/10.5209/esmp.82735

Santos, F. C. C. (2023). Artificial intelligence in automated detection of disinformation: A thematic analysis. *Journalism and Media*, *4*(2), 679–687. https://doi.org/10.3390/journalmedia4020043

Schick, N. (2020, December 22). Don't underestimate the cheapfake. *MIT Technology Review.* https://www.technologyreview.com/2020/12/22/1015442/cheapfakes-more-political-damage-2020-election-than-deepfakes/

Strauss, A. L., & Corbin, J. (2002). *Bases de la investigación cualitativa: Técnicas y procedimientos para desarrollar la teoría fundada.* Editorial Universidad de Antioquia.

Tahir, R., Batool, B., Jamshed, H., Jameel, M., Anwar, M., Ahmed, F., Zaffar, M. A., & Zaffar, M. F. (2021). Seeing is believing: Exploring perceptual differences in deepfake videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*.

Tandoc Jr., E. C., Lim, Z.-W., & Ling, R. (2018). Defining 'fake news'. *Digital Journalism, 6*(2), 137–153. https://doi.org/10.1080/21670811.2017.1360143

Torcal, M., & Magalhaes, P. C. (2022). Ideological extremism, perceived party system polarization, and support for democracy. *European Political Science Review*, *14*(2), 188–205. https://doi.org/10.1017/S1755773922000066

Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media + Society*, *6*(1). https://doi.org/10.1177/2056305120903408

Wahl-Jorgensen, K., & Carlson, M. (2021). Conjecturing fearful futures: Journalistic discourses on deepfakes. *Journalism Practice, 15*(6), 803–820. https://doi.org/10.1080/17512786.2021.1908838

Wardle, C. (2019). *First draft's essential guide to understanding information disorder*. First Draft. https://firstdraftnews.org/wp-content/uploads/2019/10/Information_Disorder_Digital_AW.pdf

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Consejo de Europa. https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c

Weikmann, T., & Lecheler, S. (2023). Cutting through the hype: Understanding the implications of deepfakes for the fact-checking actor-network. *Digital Journalism, 12*(10), 1505–1522. https://doi.org/10.1080/21670811.2023.2194665

WHO. World Health Organization. (2020). *Novel coronavirus (2019-nCoV). Situation report - 13*. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf

Wu, S. (2024). Journalists as individual users of artificial intelligence: Examining journalists' "value-motivated use" of ChatGPT and other AI tools within and without the newsroom. *Journalism*. https://doi.org/10.1177/14648849241303047

Wu, S., Tandoc, E. C., & Salmon, C. T. (2018). Journalism Reconfigured: Assessing human–machine relations and the autonomous power of automation in news production. *Journalism Studies*, *20*(10), 1440–1457. https://doi.org/10.1080/1461670X.2018.1521299