VISUAL REVIEW | Vol. 16, No. 4, 2024 | ISSN 2695-9631
International Visual Culture Review / Revista Internacional de Cultura Visual
https://doi.org/10.62161/revvisual.v16.5303

IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE TOOLS TO DETECT FAKE AND DEEPFAKE VIDEOS
The Case of Radiotelevisión Española (RTVE)

MARTA SÁNCHEZ ESPARZA [1], SANTA PALELLA-STRACUZZI [2], ÁNGEL FERNÁNDEZ FERNÁNDEZ [3]
[1] Universidad Internacional de la Empresa (UNIE), Spain
[2] EAE Business School_Madrid, Spain
[3] The Core Entertainment Science School University Centre, Spain

KEYWORDS

Fake
Deepfakes Vídeos
Artificial Intelligence
Radio Televisión Española

ABSTRACT

Concerns about the spread of false information have led media outlets to employ artificial intelligence (AI) to detect deepfakes. This research is descriptive-exploratory, a literature review and interviews were conducted. It reveals the transformative impact of AI by highlighting its use to verify the authenticity of content. In this area, RTVE combines traditional methodologies with others based on AI, and leads the development of several tools in collaboration with several universities. These tools have already yielded satisfactory results in the detection of these materials, strengthening the veracity of the information and increasing public confidence in their contents.

Received: 17 / 04 / 2024
Accepted: 05 / 06 / 2024

1. Introduction

Disinformation, understood as the deliberate and usually covert dissemination of false information with the aim of manipulating public opinion or destabilising audiences (Fernández, et al., 2020), has undergone a very significant transformation in the digital age. Although lies and propaganda have been used throughout history as tools of political and social control (Pineda, 2004), it is clear that never before has humanity had such powerful and effective tools for spreading hoaxes, lies and other false content. Every historical period has used its own technology to spread false information and propaganda. Today, however, the development of information technologies has completely changed this landscape, allowing an unprecedented amount of disinformation to be disseminated. The proliferation of the internet and the increased participation of users in the creation, classification and distribution of all types of digital content has meant that our daily activities have become a constant exchange of information and data (Fernández, 2017). While most of this content does not pose a problem of veracity, it is increasingly common for users to choose to share completely or partially false content that reinforces their own biases and opinions (Olmo and Romero, 2019).

Undoubtedly, one of the most worrying consequences of this situation is the devaluation of the concept of truth itself. Although it is generally possible to interpret this phenomenon as a consequence of the relativism that characterises postmodern thought, Kavanagh and Rich (2018) believe that behind it lie social trends that are increasingly rooted in today's digital culture, such as the growing disagreement between the subjective interpretation of facts and the objective analysis of the data that support them, the growing influence of opinion and personal experience on our interpretation of reality, and the deepening distrust of media and sources of objective information that were previously considered legitimate.

In this context, the role of images is particularly critical. Because of their capacity for symbolic projection, images represent a highly emotive model of communication, making them a privileged medium for the dissemination of false content and disinformation (Hameleers et al., 2020). Images are particularly useful in capturing the attention of audiences overwhelmed by information overload and in lending credibility and believability to false information. Particularly on social networks, content that includes images tends to spread more than content that is purely textual. For example, tweets that include photos or videos receive approximately 18% more clicks, 89% more likes, and 150% more retweets than those that are text-only (Cao et al., 2020).

The ability of AI to create deepfake videos has profound implications for the very structure of digital media. While the ultimate capabilities of AI have yet to be realised and there is little regulation, the intersection of digital media and artificial intelligence is a central element in the evolution of media communication today. Deepfakes and other types of content generated by AI systems blur the lines between what is real and what is fake, undermining the public's trust in the information they receive. With deepfakes, the ability to distort reality has undergone a radical transformation. For this reason, it is crucial to examine this intersection from a variety of perspectives.

In this context, it is relevant to analyse how a public media outlet such as Spanish Radio Television (Radiotelevisión Española, RTVE) deals with the proliferation of this disinformative content and how it uses new tools based on artificial intelligence to detect these fake videos and deepfakes.

Radiotelevision Española, a public broadcaster with 6,600 employees, is currently undergoing a major digital transformation. In this process, artificial intelligence is the main catalyst for change. One of the main applications of this technology is the analysis of content and the detection of fake images and audio using AI.

2. Objectives and Methodology

The general objective of this study is to describe the implementation of Artificial Intelligence (AI) tools in Radiotelevisión Española (RTVE) and how they are used to detect fake and deepfake videos. By implementation, we mean the incorporation of these tools into the company's work processes.

To this end, the following specific objectives will be pursued:

- To describe how AI has been implemented in RTVE.
- To get to know the AI tools used in different areas of RTVE.
- To explore the use of AI tools in the detection of deepfakes in RTVE.

From a methodological point of view, a literature review and an exploratory-descriptive field research were carried out, studying the case of Radiotelevisión Española (RTVE). For the fieldwork, the interview technique was used, which allowed the information to be collected in order to respond to the objectives.

The instrument used was a structured interview script. The script was constructed with 11 different questions according to the role played by the interviewee within RTVE. The validity of the interview script was verified through the judgement of three experts, who confirmed that the questions responded to the objectives set (Palella and Martins, 2017:106).

In this regard, we interviewed Urbano García, director of Innovation and Digital at RTVE, Pere Vila, director of Technological Strategy at RTVE, and Borja Díaz-Merry, director of the Verification Service at RTVE (Verifica RTVE). The interviews took place between 17 and 22 November.

The specific sample of RTVE professionals was selected based on the level of contribution of relevant information on the implementation of Artificial Intelligence (AI) tools in a media company and, more specifically, on the use of these new tools in the detection of videos, deepfakes and fake videos in the case of Radiotelevisión Española (RTVE).

3. Literature Review

3.1. Artificial Intelligence in the Media

Artificial Intelligence (AI) is a branch of computer science that focuses on creating intelligent machines capable of performing tasks previously performed by humans. It is increasingly being implemented in the media industry, leading to significant changes and challenges.

AI is used in media to optimise and improve operations, such as data analysis and multimedia content generation (Sančanin & Penjišević, 2022). It can also be used to automate processes, including social media management, where algorithms can be trained to analyse users' actions, preferences and reactions (Al Husseiny, 2023).

AI is also being used in the news industry to disrupt traditional approaches, using machine learning to plan, schedule and optimise processes that are becoming increasingly sophisticated (Kalinová, 2022). The implementation of AI on social media platforms is becoming inevitable, with applications such as chatbots that detect harmful behaviour, analyse data and strategies (De Lima-Santos & Wilson, 2022). It can be argued that AI has the potential to transform businesses in this sector and their functions (Sadiku et al 2021).

This transformation frees media professionals from routine tasks and allows them to produce higher quality content. However, it also raises concerns about the growing dependence on technological platforms and the threat to editorial independence. Media workers perceive a threat to their jobs and a potential loss of their symbolic capital as intermediaries between reality and audiences (Peña-Fernández et al., 2023).

The implementation of AI in the media therefore poses social and epistemological challenges for journalists and the profession. There is also a debate in the European Union about the use of these technologies in the media. In this area, regulatory frameworks on AI rarely include the media; when they do, they address issues such as disinformation, data, AI literacy, diversity, plurality and social responsibility (Porlezza, 2023).

The European Union is also waging a battle against disinformation content, launching various strategic plans and setting up working groups. According to the European Commission, disinformation is not just a side issue, but an ecosystem (Jerónimo and Sánchez-Esparza, 2022).

To deal with the content generated in such an ecosystem, journalists rely on a combination of traditional and digital methods. A study by Haidar (2023) highlights the reliance of reporters on free websites and tools (69.2%). If the importance of AI in journalism and media in analysing and creating content is undeniable, its role in verifying information is crucial.

3.2. Fake and Deepfake Videos. Similarities and Differences

Fake videos are videos that have been manipulated or generated using artificial intelligence technology to create content that is not real or accurate. Several techniques have been developed to detect these

fake videos, such as methods based on multimodal learning that combine audio, video and physiological information (Stefanov et al., 2022). Biometric-based forensic techniques have also been used. These techniques aim to identify discrepancies or anomalies in videos that indicate tampering or forgery (Matthews, 2023; Timoth and Shih, 2011;).

Among the most dangerous types of disinformation content are manipulated videos and, even more so, deepfakes, in which one person's face is superimposed on another person's body to create false and convincing content. Detecting these deepfakes is a major challenge and is becoming increasingly difficult due to rapid advances in facial manipulation techniques (Al-Khazraji et al., 2023).

In its most common definition, deepfakes include photos, videos and audio digitally generated using artificial intelligence techniques (Bañuelos, 2022) that realistically represent individuals performing actions or expressing words that they never performed or said (Cerdán and Padilla, 2019). It is therefore content that is explicitly designed to generate false and misleading information. Although, as we have just indicated, the term can be applied to different formats, its use has become increasingly specific and mainly refers to digitally created videos. In these videos, a person's face and/or voice is superimposed on content previously recorded by another person or merged with an image digitally generated using machine learning and deep learning techniques.

It is important to note this nuance, because although most deepfake videos consist of the superimposition of one person's face on another person's body, this category also includes entirely new images and sounds generated directly by AI systems through the synthesis of large datasets, without necessarily starting from a previous real image or sound (Karnouskos, 2020).

Although deepfakes are powered by artificial intelligence systems based on complex technology, they can often be created using easily accessible tools and services available to the general public. In fact, most of the tools currently used to create deepfakes have low technical requirements and can be easily used on standard home computers with mid-range graphics cards.

The accessibility of these technologies and the low learning curve of the tools most used to create deepfakes is one of the greatest risks posed by this type of content. The fact that users with little technical knowledge can create extremely realistic fake images encourages the creation and distribution of this type of content, which logically increases the risks posed by its use. When we consider the ease with which deepfakes can be distributed through social networks, the serious implications for the spread of hoaxes and other forms of disinformation become clear.

However, tools that are also generated with artificial intelligence and that aim to detect these fake videos have emerged. Thus, in addition to the methodologies developed by some experts from the perspective of journalists (Sohrawardi et al., 2019), work has been added on the effectiveness of neural networks (Shilma et al., 2023). These tools can distinguish between real and fake faces thanks to training models (Haseena et al., 2023).

Fake videos and deepfakes share similarities in that both involve the creation of manipulated content and can be used to alter the appearance or actions of individuals, leading to potential disinformation and distortion of the truth (Matthews, 2023).

Deepfakes pose unique challenges as they can create a high degree of epistemic risk, which could lead to scepticism about the knowledge of online videos (Liz-López, 2023). In general, deepfakes represent a more advanced and sophisticated form of manipulation and require a much greater investment in techniques and processing time (Díaz-Merry, B., interview on 22.11.2023).

4. Results

4.1. Implementation of Artificial Intelligence in RTVE

The Spanish Radio and Television Corporation (Corporación Pública de Radio y Televisión Española. RTVE) has been integrating Artificial Intelligence (AI) technologies into different processes and departments for years. As early as 2015, it launched a research programme on the possibilities offered by intelligent information processing systems (Aramburu et al., 2023). This programme brings together experts in AI processes, students and professors, and is supported by initiatives such as the RTVE-UAB Chair (with the Autonomous University of Barcelona) and the Observatory for News Innovation in the Digital Society (OI2).

The arrival in 2021 of José Manuel Pérez Tornero, Professor of Journalism at the Autonomous University of Barcelona, who has decided to create the Directorate of Innovation and Digital, headed by

journalist Urbano García, can also be considered a key date. This is an area that aims to implement a strategic plan to transform a television station that produces some digital content into a "fully digital core company whose activities include television" (García, U., personal communication, 17 November 2023).

This new Innovation and Digital Directorate has absorbed the Observatory for Innovation (OI2), as well as the management of the chairs and the Audiovisual Innovation Laboratory (Lab) of RTVE, a department dedicated to exploring new narratives. This Directorate is also in charge of the strategy related to new media and the transition of the entire company to the new digital model, reconciling the values that make up the mission of public television with the use of new technologies such as AI for content development.

In this transition process, AI-based technologies are already having an impact on the way television is produced and on the business model itself. Faced with these challenges, the directors of RTVE, together with other public companies such as the State Industrial Holding Company (Sociedad Estatal de Participaciones Industriales. SEPI), have undertaken a general reflection on the limits, risks and opportunities involved.

According to Pere Vila, RTVE's Director of Technological Strategy, artificial intelligence will infiltrate all of the company's activities, not only in areas such as documentation - through the metadata of all of RTVE's archives - but also in content analysis and automation projects, image processing and colouring, voice cloning, avatar generation, selection of people, relationships with the audience, and content recommendation, among others (Vila, P., personal communication, 22 November 2023).

RTVE is currently working with AI technologies in the areas shown in the table below:

Table 1. Areas of Artificial Intelligence technologies in RTV

Content analysis	Content generation	Other applications
- Speech recognition technologies are used to generate automatic subtitles in real time during live broadcasts. - Artificial intelligence is used for automatic indexing of the document archive, allowing information to be organised and labelled more efficiently and accurately. - Another use is the analysis and counting of topics covered or video content, which is useful for Corporate Social Responsibility reporting or, for example, to quantify the time in which sign language is used. - Recommendation systems are used based on users' interests, offering content that matches their preferences.	- Pre-processed data and information are used to generate text, graphics and audio automatically. - Creation of artificially intelligent voices that can speak naturally, as if it were a human voice.	- Increasing the quality of archival images by removing noise, colouring them and improving their sharpness. - Technology applied to projects against disinformation and new forms of verification. - Creation of personalised avatars, created entirely with artificial intelligence.

Source: Own elaboration.

4.2. AI Tools in Different Areas of RTVE

In general, the tools used are external and are acquired through the purchase of licences. Once acquired, the company's technology team carries out tests to see how these applications can be used in RTVE's daily work.

The following table shows the main tools based on AI technologies used at RTVE:

Table 2. AI tools in RTVE

AI tools	Use
Lexica	AI-generated image creation and image detection
Stable Diffusion	Creating high quality images from text
ChatGPT	Generation of texts and other content.
Dall-E	Image creation
HeyGen	Avatar generation
Studio D-ID	Avatar generation
Eleven Labs	Voice cloning
Runway and Stable Diffusion	Transformation of images and video clips
Adobe (functionalities)	Generative filling
Open Access Toolbox	Verification of information and deepfakes
IVERES project tools	Transcription and translation, detection of fake audios and videos

Source: Own elaboration.

4.2.1 Technology Behind AI Tools Applied to RTVE's Areas of Activity

Behind the tools used to detect fake videos and deepfakes are the following technologies

- Face and voice recognition: When analysing the characteristics of a face using facial recognition algorithms, the analysed image is compared with images of known faces stored in a database, looking for inconsistencies (Guarnera & Battiato, 2023). In the case of voice recognition, timbre, pitch and pronunciation are analysed and compared with voices stored in databases.
- Metadata analysis: Programmes for modifying images, videos and sound bites add their own metadata to the file or modify existing metadata (date and time of creation or modification, software used, location, etc.). Again, inconsistent information can reveal a deepfake.
- Forensic analysis: Looking for inconsistencies in light, shadows, reflections and small disturbances at the pixel level. In the case of images, techniques such as Error Level Analysis (ELA) (Martín-Rodríguez et al., 2023) and systems such as ProtoExplorer (Bouter et al., 2023) are used. In the case of video, the analysis also includes movement and perspective or facial expressions. In the case of audio, the objects of analysis are usually the waveform and the spectrogram (frequency range, harmonics, background noise, etc.).
- Machine Learning and Artificial Intelligence: Just as a neural network can be trained to generate multimedia material, it can also be trained to recognise the synthetic generation of the same material. That is, a neural network can be used to detect the intervention of another neural network. As with other deep neural networks, training these systems requires large datasets, including real and manipulated images.
- Blockchain technology: As in any other field where it is necessary to guarantee the impossibility of subsequent modification of a transaction, this technology can be used to generate and guarantee the authenticity of a multimedia element. The storage of such elements in a blockchain system makes it virtually impossible to alter or manipulate them without detection. Blockchain can guarantee the origin and traceability of data, creating a secure platform for storing and exchanging multimedia information (Rashmi et al., 2023).

4.3. Instances of the Use of Artificial Intelligence in Content Creation

Among the products created at RTVE thanks to the use of AI is the RTVEIA project (www.rtveia.es), a website from which 70,000 news items with text, images and synthetic voices were launched on the afternoon of the last general election in 2023, reporting the results in real time in almost 5,000 Spanish municipalities with less than 1,000 inhabitants. According to Urbano García, this type of content reinforces RTVE's role as a public service by structuring the territory in places where there is no media presence. There, AI makes it possible to provide a better news service without replacing journalists. The project, with the participation of the company Narrativa, Monoceros Labs, the University of Castilla-La Mancha, the University of Granada, ONCE, AWS and RTVE's Castilla-La Mancha Territorial Centre, received the IBC 2023 Award for Innovation and Social Impact in September 2023.

Another product developed by RTVE using AI is RTVE2030 (www.rtve2030.rtve.es), a website that analyses the content of current affairs programmes every day and measures the time devoted to each of the Sustainable Development Goals (SDGs). These analyses (Figure 1) are later used to produce reports on the company's CSR policies.

Figure 1. RTVE2030 website, with analytics on SDGs using AI

Source: www.rtve2030.rtve.es

A third project to be launched in February 2023 is Hiperia, an artificially intelligent avatar who will present a weekly music and youth culture slot for Radio 3. The character was created thanks to collaboration between Radio 3 and the Technology Strategy, Innovation and Digital and Graphics departments. Both the character and his voice, the script and the content of the programme are created using artificial intelligence (Vila, P., personal conversation, 22.11.2023).

Figure 2. Image of Hiperia, an AI-created presenter for a Radio 3 slot. Source:

www.rtve.es/radio

4.4. The Use of AI in the Detection of Fake and Deepfake Videos at RTVE

4.4.1 The Verifica RTVE Service

A third project to be launched in February 2023 is Hiperia, an artificially intelligent avatar who will present a weekly music and youth culture slot for Radio 3. The character was created thanks to collaboration between Radio 3 and the Technology Strategy, Innovation and Digital and Graphics departments. Both the character and his voice, the script and the content of the programme are created using artificial intelligence (Vila, P., personal conversation, 22.11.2023).

The Verifica RTVE team currently consists of six journalists, all with degrees in Information Science, who have been trained to carry out information verification and research tasks through specific training within RTVE. These professionals are dedicated to verifying the political discourse around major events, such as parliamentary debates or debates during an election campaign, in coordination with journalists from the news services.

The journalists of Verifica RTVE work proactively in the monitoring of social networks, looking for suspicious, fraudulent, false or misleading news. On a second level, they work internally at the request of RTVE's news departments, which need to check the veracity of videos, photos, reports and documents before using them in the information that will be broadcast that day.

According to the head of Verifica RTVE, Borja Díaz-Merry, it is this line of internal work that is growing faster and requires more effort from the team, especially since the outbreak of the war in Ukraine in February. If, before the conflict, the internal demand for verification of material from the news services was two or three per month, since the Russian invasion the rate of requests has risen to two or three per day, both for the two editions of the News Program (Telediario) and for RTVE's 24-Hour channel.

From that moment on, the internal verification dimension took on a prominent role in Verifica RTVE's work, checking videos on demand for the news programmes, but also for RTVE's territorial services, where videos of floods, police operations, events, etc. of all kinds are received. The team continues to analyse and verify false content that goes viral on social networks, but the bulk of its work has shifted to ensuring the reliability of the public broadcaster's news programmes.

In this context, the service usually detects three to four stories a day that go viral on social networks, some of which are analysed according to journalistic criteria, and devotes more time to these stories as they do not have to be published during the day. On the other hand, requests from news services must be dealt with more quickly. According to Díaz-Merry, the volume of checks they carry out at the internal request of news services is around twenty per month.

It is important to distinguish between fake videos - known as shallow fakes - and deepfakes. While the former are videos manipulated by simple editing, deepfakes are more sophisticated and elaborate. They are digital creations in which a real recording is superimposed over millions of photographs to impersonate a person's face. In some cases, the same technology is used to impersonate a voice.

According to Díaz-Merry, this level of sophistication means that deepfakes are not as common as simple fake and manipulated videos. In fact, while Verifica RTVE usually detects two or three shallow fakes per month, deepfakes do not appear as often. Not all these videos are verified, as RTVE practices "responsible verification", which means that it examines the danger of the content and the extent to which it has gone viral and decides whether it is better to publish its verification or simply warn the news services to prevent further dissemination of the issue.

4.4.2. AI Tools in the Detection of Fake and Deepfake Videos in RTVE

To analyse these fake videos, professionals use AI tools combined with journalistic analysis. Most of these tools are publicly available and free and are offered to users in Verifica RTVE's so-called "toolboxes", which can be found on its website. There are two toolboxes, a basic one and an advanced one, which anyone can use to check false information.

On the other hand, RTVE, together with the Autonomous University of Barcelona, is leading the IVERES (Investigation, Verification and Response) project, with European Next Generation funding, in which the University of Granada, the Polytechnic University of Barcelona and the Carlos III University are also involved, and which is developing specific toolboxes for Verifica RTVE using artificial intelligence. There are three types of tools: transcription and archiving tools, fake audio detection tools and fake and deepfake video detection tools. Each is being developed by one of the three participating universities.

Within the IVERES project, the tool developed and supervised by the Carlos III University for the archiving of material and transcription in several languages is used by the journalists of Verifica RTVE with very satisfactory results (Díaz-Merry, B., personal interview, 22.11.2023). In fact, it has already been used to verify dialogue in long videos about the war in Ukraine and the Taliban's seizure of power in Afghanistan, with transcriptions and translations from Persian or Pashto into Spanish and English.

The second of the tools developed in the IVERES project aims to detect false audio through an artificial intelligence model based on the training of neural networks with a database of voices. The tool is being developed at the University of Granada and detects voices generated by artificial intelligence using artificial intelligence technologies. Although it is not yet available for everyday use by journalists, they can make specific requests to the tool's development team.

The same applies to the tool for detecting fakes and deepfakes being developed by the Polytechnic University of Barcelona. Journalists report on examples where the tool has been used and the team of

developers tests the tool's reliability, although it is not yet available. It is precisely when it comes to verifying videos with fake voices and images that professionals are currently encountering the greatest difficulties, which is why there are great expectations for these two tools (Díaz-Merry, B., personal conversation 22.11.2023).

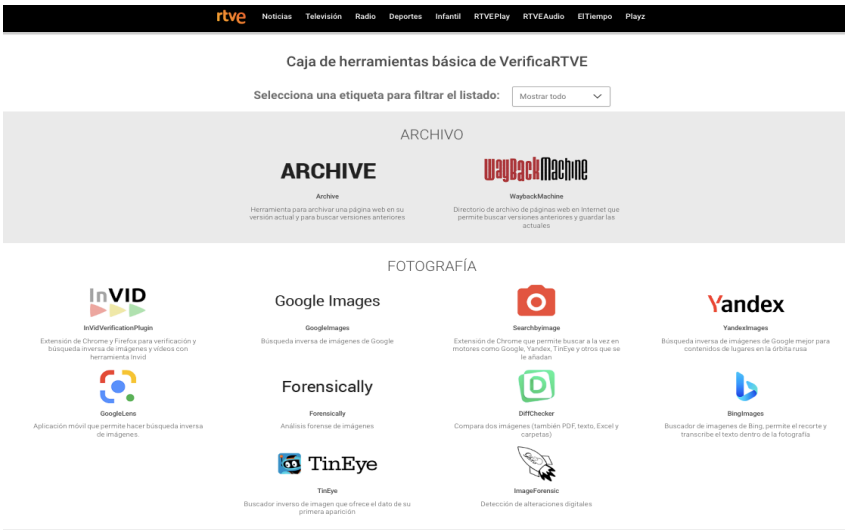
While these last two tools are arriving, Verifica RTVE uses frame-by-frame video analysis and free access tools for professional verifiers, such as INVID We Verify, developed by the Agence France-Presse (AFP). There are also open access tools for video analysis, powered by artificial intelligence, that enable reverse searches from search engines such as Google images, Yandex or Bing, among others. Traditional search engines perform text-to-text or text-to-image queries. Reverse image search is a technique that goes further and is used to find identical or similar images based on a given query image. This technique is commonly used in image search engines that have large databases of uploaded images (Nandini et al., 2022).

In Verifica RTVE, Microsoft Azure was also used in terms of video analysis tools, although its performance is not as good, and it is starting to be replaced by other tools.

Finally, Díaz-Merry stresses that, in addition to the support of all these tools, when it comes to verifying deepfakes, practitioners rely mainly on frame-by-frame human analysis, which cannot yet be replaced by AI tools.

The greatest proliferation of fake videos and deepfakes occurs in emergency contexts such as health crises, wars or disasters, although verifiers report that the main international trend is for these videos to be fabricated to destroy the reputation of women through sexual content. Propagandistic narratives are also produced from countries in conflict, as in the case of the war in Ukraine and the Israeli-Palestinian war, sometimes using fake images produced by artificial intelligence. Verifica RTVE professionals have found examples of disinformative audiovisual content in all these scenarios. These professionals are aware of the need for continuous training in the use of new tools and artificial intelligence, in order to be able to face the challenges that AI itself creates in the form of disinformative content.

Figure 3. Basic toolbox of Verifica RTVE



Source: www.rtve.es/noticias/verificartve

5. Discussion and Conclusions

At RTVE, AI is being implemented with a global and strategic vision, and its managers intend to infiltrate all areas of the company in a short time (Vila, P., personal conversation, 22-22-2023), affecting a series of tasks that include optimising operations, analysing data and generating new content, in line with what has been stated by authors such as Sančanin & Penjišević, (2022). Automating processes and interacting with audiences by training algorithms is already improving business efficiency, according to Al Husseiny (2023).

In the specific case of RTVE, AI tools are used to analyse content and create new content, such as text, images, synthetic voices and avatars. In this sense, tools such as Lexica, ChatGPT, Dall-E, HeyGen, Studio D-ID, Eleven Labs, Runway and Stable Diffusion are used, as well as Adobe's functionalities. The list of these tools is constantly growing as different projects are developed.

RTVE's mission is to "provide and guarantee the public service of radio and television owned by the State" (www.rtve.es/corporacion/quienes-somos) and, in accordance with this mandate, it must ensure the correct transmission of information, truthful information to which all Spanish citizens are entitled, as set out in Article 20 of the Spanish Constitution. In this context, the creation of bodies such as Verifica RTVE, which is specifically dedicated to combating disinformative content and preventing it from being erroneously broadcast in the public broadcaster's programming, makes sense.

To deal with disinformative content, RTVE journalists use a combination of traditional and digital methods, relying mostly on free tools, together with others acquired by the company itself, in line with the study carried out by Haidar (2023).

To verify the information, RTVE uses traditional tools such as frame-by-frame analysis, along with several other free and paid digital tools, many of which are based on AI technologies. This is the case of the reverse image search technologies provided by several search engines, which are offered free of charge in the verification toolbox available on the Verifica RTVE website.

If detecting fake videos can be a complex task, even more so is the identification of deepfakes, which RTVE faces as a major challenge due to the increasing complexity of facial manipulation techniques, in line with what Al-Khazraji et al. (2023) stated. Particularly complex is the verification of fake videos shared via WhatsApp and deepfakes, where there are no original voices to contrast with the fake ones (Díaz-Merry, B., personal conversation, 22-22-2023).

In this context, RTVE is working on the development of its own AI-based tools in collaboration with several universities, as part of the IVERES project, where a new tool for the verification of fake audio - through training based on voice databases - and another for the detection of fake and deepfake videos are currently in the testing phase.

The new tools aim to achieve greater efficiency in the detection and deactivation of this content, allowing RTVE to fulfil its role as a guarantor of the veracity of information and to guarantee the quality and reliability of the content broadcast by the channel, thereby increasing public confidence.

6. Acknowledgements

This text is part of the Artificial Intelligence and New Frontiers in Communication research project of the Universidad Internacional de la Empresa (UNIE) Madrid, Spain.

We would like to thank the people at RTVE for their collaboration and generosity in providing us with information.

References

Al Husseiny, F. (2023). The Rising Trend of Artificial Intelligence in Social Media: Applications, Challenges, and Opportunities. In S. Kaddoura (Ed.), *Handbook of Research on AI Methods and Applications in Computer Engineering* (pp. 42-61). IGI Global. https://doi.org/10.4018/978-1-6684-6937-8.ch003

Al-Khazraji, S. H., Saleh, H. H., Khalid, A. I. & Mishkhal, I. A. (2023). Impact of Deepfake Technology on Social Media: Detection, Misinformation and Societal Implications. *The Eurasia Proceedings of Science Technology Engineering and Mathematics, 23*, 429-441. https://doi.org/10.55549/epstem.1371792

Aramburú, L. G., López, I. & López, A. (2023) Inteligencia artificial en RTVE al servicio de la España vacía. Proyecto de cobertura informativa con redacción automatizada para las elecciones municipales de 2023. *Revista Latina de Comunicación Social*, *81*, 1–16. https://doi.org/10.4185/rlcs-2023-1550

Bañuelos, J. (2022). Evolución del Deepfake: campos semánticos y géneros discursivos (2017-2021*). Revista ICONO 14. Revista Científica De Comunicación Y Tecnologías Emergentes, 20*(1). https://doi.org/10.7195/ri14.v20i1.1773

Bouter, M. D. L. D., Pardo, J. L., Geradts, Z., & Worring, M. (2023). ProtoExplorer: Interpretable Forensic Analysis of Deepfake Videos using Prototype Exploration and Refinement. *arXiv (Cornell University)* https://doi.org/10.48550/arXiv.2309.11155

Cao, J., Qi, P., Sheng, Q., Yang, T., Guo, J., & Li, J. (2020). Exploring the Role of Visual Content in Fake News Detection. In K. Shu, S. Wang, D. Lee y H. Liu (Eds.). *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities*, 141-161. Springer International Publishihaidng. https://doi.org/10.48550/arXiv.2003.05096

Cerdán, V. & Padilla, G. (2019). Historia del fake audiovisual: deepfake y la mujer en un imaginario falsificado y perverso. *Historia y comunicación social, 24*(2), 505-520. https://dx.doi.org/10.5209/hics.66293

De Lima-Santos, M.-F., & Wilson C. (2022). Artificial Intelligence in News Media: Current Perceptions and Future Outlook. *Journalism and Media*, *3*(1), 13-26-. https://doi.org/10.3390/journalmedia3010002

Fernández, A. (2017). Relatos híbridos: El papel de la narratividad en la visualización de información interactiva [Tesis doctoral, Universidad Europea]. Repositorio Abacus https://193.147.239.238/handle/11268/6981

Fernández, A., Revilla, A. & Andaluz, L. (2020). Análisis de la caracterización discursiva de los relatos migratorios en Twitter. El caso Aquarius. Revista Latina de Comunicación Social, (77), 1-18. https://doi.org/10.4185/RLCS-2020-1446

Guarnera, L., & Battiato, S. (2023). An Overview of Deepfake Technologies: from Creation to Detection in Forensics.

Haidar, H. (2023). Using artificial intelligence to verify media content on the Internet. A survey study of journalists working in Iraqi media institutions. *International Journal of Media Studies and Communication Sciences*. https://doi.org/10.36772/arid.aijmscs.2023.485

Hameleers, M., Powell, T. E., Van Der Meer, T. G., & Bos, L. (2020). A Picture Paints a Thousand Lies? The Effects and Mechanisms of Multimodal Disinformation and Rebuttals Disseminated via Social Media. *Political Communication, 37*(2), 281-301. https://doi.org/10.1080/10584609.2019.1674979

Haseena, S., Saroja, S., & Nivetha, A. (2023). TVN: Detect Deepfakes Images using Texture Variation Network. *Inteligencia artificial, 26*(72), 1-14. https://doi.org/10.4114/intartif.vol26iss72pp1-14

Jankowicz, N., Hunchak, J., Pavliuc, A., Davies, C., Pierson, S., & Kaufmann, Z. (2021) Malign Creativity: How Gender, Sex and Lies Are Weaponized Against Women Online, Washington, D.C.: Wilson Center. https://www.wilsoncenter.org/publication/malign-creativity-how-gender-sex-and-lies-are-weaponized-against-women-online

Jerónimo, P., & Esparza, M. S. (2022). Disinformation at a Local Level: An Emerging Discussion. *Publications, 10*(2), 15. https://doi.org/10.3390/publications10020015

Kalinová, E. (2022). Usage of artificial intelligence on social media in europe. *Ad Alta, 12*(2), 330-333. https://doi.org/10.33543/1202330333

Karnouskos, S. (2020). Artificial intelligence in digital media: The era of deepfakes. *IEEE Transactions on Technology and Society, 1*(3), 138-147. https://doi.org/10.1109/tts.2020.3001312

Kavanagh, J. & Rich, M. D. (2018). *Truth decay: An initial exploration of the diminishing role of facts and analysis in American public life*. Rand Corporation.

Liz-López, H. ; Keita M. , Taleb-Ahmed, A. Abdenour H. , Huertas-Tato, J., & Camacho D. (2023). Generación y detección de contenidos audiovisuales multimodales manipulados: Avances, tendencias y desafíos abiertos. *Fusión de Información*, pp.102-103.

Martin-Rodriguez, F., Garcia-Mojon, R. & Fernandez-Barciela, M. (2023). Detection of AI-Created Images Using Pixel-Wise Feature Extraction and Convolutional Neural Networks. *Sensors, 23*(22). http://dx.doi.org/10.3390/s23229037.

Matthews, T. (2023). Deepfakes, fake barns, and knowledge from videos. *Synthese*, *201*(2). https://doi.org/10.1007/s11229-022-04033-x

Nandini S, Akshay B G, Brunda A N, Chandana A M, & Divyashree S R. (2022). Advanced reverse image search and profile creation using machine learning. *International Journal of Advanced Research in Science, Communication and Technology*, 586–589. https://doi.org/10.48175/ijarsct-5417

Olmo, J. & Romero, A. (2019). Desinformación: Concepto y perspectivas. Análisis del Real Instituto Elcano (ARI), (41). https://www.realinstitutoelcano.org/analisis/desinformacion-concepto-y-perspectivas/

Palella, S y Martins, F. (2017). *Metodología de la investigación cuantitativa*. FEDEUPEL

Peña-Fernández, S., Meso-Ayerdi, K., Larrondo-Ureta, A., & Díaz-Noci, J. (2023). Without journalists, there is no journalism: the social dimension of generative artificial intelligence in the media. el Profesional de la Información. https://doi.org/10.3145/epi.2023.mar.27

Pineda, A. (2004). Más allá de la historia: aproximación a los elementos teóricos de la propaganda de guerra. En A. Pena (Ed.), *Comunicación y guerra en la historia* (pp. 807-823). Santiago de Compostela: Tórculo. http://hdl.handle.net/11441/64448

Porlezza, C. (2023). Promoting responsible AI: A European perspective on the governance of artificial intelligence in media and journalism. *Communications, 48*(3), 370-394. https://doi.org/10.1515/commun-2022-0091

Rashmi, C., Bhargavi, V., Samhitha, S., Anjana, Y., & Saivaishnavi, V. (2023). Fake detect: a deep learning ensemble model for fake news detection (ml). Turkish Journal of Computer and Mathematics Education (TURCOMAT), 14(03), 684-688.

Sadiku, M. N. O., Ashaolu, T. J., Ajayi-Majebi, A., & Musa, S. M. (2021). Artificial Intelligence in Social Media. International Journal Of Scientific Advances, 2(1). https://doi.org/10.51542/ijscia.v2i1.4

Sančanin, B., & Penjišević, A. (2022). Use of artificial intelligence for the generation of media content. Social Informatics Journal, 1(1), 1-7. https://doi.org/10.58898/sij.v1i1.01-07

Shilpa, B., Kamath, A., Bhat, H., & Sathwik A M. (2023). Unmasking deepfakes: Using Resnext and LSTM to detect deepfake videos. *International Journal of Advanced Research in Science, Communication and Technology*, 524–528. https://doi.org/10.48175/ijarsct-8639

Sohrawardi, S., Chintha, A., Thai, B., Seng, S., Hickerson, A., Ptucha, R. & Wright, M. (2019). Póster: Hacia una detección sólida de deepfakes en mundo abierto. Actas de la Conferencia ACM SIGSAC de 2019 sobre seguridad informática y de las comunicaciones. https://doi.org/10.1145/3319535.3363269

Stefanov, K., Paliwal, B., & Dhall, A. (2022). Visual Representations of Physiological Signals for Fake Video Detection. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2207.08380

Timothy, K., & Shih. A. (2011). Video Forgery. *2011 14th International Conference on Network-Based Information Systems*.

Vedamurthy, H. K., Ravi, & Gururaj. (2022). A reliable solution to detect deepfakes using Deep Learning. *2022 Fourth International Conference on Cognitive Computing and Information Processing (CCIP)*.