Journal of Social Knowledge Education (JSKE)
Vol. 6, No. 3, August 2025, pp. 386~399
ISSN: 2722-046X, DOI: 10.37251/jske.v6i3.1798

Deepfake: A Study on Knowledge of Media Practitioners in Cotabato Province, Philippines

Aravila Quindo Demon[1] and Vilma M. Santos[2]
[1,2]College of Arts and Social Sciences, University of Southern Mindanao, Kabacan, Cotabato, Philippines

Article Info

Article history:

Received May 31, 2025
Revised Jun 20, 2025
Accepted Jul 29, 2025
Online First Aug 31, 2025

Keywords:

Artificial Intelligence
Deepfake
Knowledge
Media Practicioners

ABSTRACT

Purpose of the study: Deepfake technology uses an Artificial Intelligence algorithm to convincingly manipulate images, videos, and audio thereby replacing individual's likeness with that of another. The primary objective of the study is to assess the level of knowledge regarding deepfakes of media practitioners in Cotabato Province. Additionally, it examines the relationship between the practitioners' socio-demographic characteristics, such as their years of experience in the media field and the undergraduate academic program taken, and their level of knowledge on deepfake.

Methodology: The study employed a descriptive-correlational research design to examine the relationship between selected variables. A cluster random sampling technique was used to determine the sample. The second district of Cotabato Province comprises 15 radio stations, from which 9 stations were randomly selected as representative clusters. A total of 25 media practitioners participated in the study, including news writers, DJs, reporters, and broadcasters. Technicians were intentionally excluded from the sample, as the focus was on individuals directly involved in the production and dissemination of news and information.

Main Findings: Findings revealed that media practitioners demonstrated good knowledge of deepfake content, its creation, and the software commonly used for generating deepfakes. However, their knowledge was limited when it came to deepfake detection and the software tools available for identifying such manipulated content. Furthermore, a significant relationship was found between knowledge of deepfake content and the years of experience as a media practitioner. In contrast, no significant correlation was observed between years of experience and knowledge of deepfake creation, detection, or the corresponding software used for either process.

Novelty/Originality of this study: Years of experience in media practice correlate positively with deepfake content knowledge, but not with knowledge of detection or creation tools, suggesting that experiential exposure does not necessarily equate to technical proficiency.

Corresponding Author:
Aravila Quindo Demon
College of Arts and Social Sciences, University of Southern Mindanao
USM Avenue, North Cotabato, Kabacan, Cotabato, Region XII, Philippines, 9407
Email: rktagare@usm.edu.ph

1. INTRODUCTION

The proliferation of falsehoods, fantasies, and fake news has long accompanied the evolution of communication and journalism, highlighting an urgent need to examine how these trends challenge the practice and societal role of media [1]. Paunovic [2] argues that media practitioners must gather and share accurate information, maintain truthfulness and neutrality, and use responsible communication, given their power to incite or alleviate tensions. The primary responsibility of media practitioners is to avoid harm by ensuring that information does not spark conflict or damage reputations. Upholding impartiality, avoiding manipulation, and offering balanced analysis are essential to building trust.

In the digital era, artificial intelligence increasingly influences the production and distribution of information. One of its damaging results is deepfake technology. Deepfakes falsely manipulate media by closely mimicking real individuals and events. Recently, deepfakes pose a substantial global security and media threat. For example, 37% of organizations report incidents of deepfake voice fraud, and 29% were victimized by deepfake videos [3]. Despite this increasing threat, the public is still not well-informed about deepfakes [4]. Predominantly, only a few communication studies have addressed deepfakes [5]. Papers using qualitative or quantitative methods such as interviews and questionnaires in their research findings are scarce [6]. This data indicates a research gap that requires further exploration.

This research highlights the persistent need to understand and mitigate deepfakes' harmful effects. Rapid technological progress adds to this urgency. It aims to enhance public awareness and increase precaution in society regarding these impending threats. Exposure to deepfakes intensifies vulnerability to manipulation, exploitation, defamation, and fraud. The viral nature of such content worsens these apprehensions. It can destabilize public trust and threaten the integrity of democratic processes [7]. Thus, a crucial first step in lessening these hazards is the ability to distinguish real content from altered material. In this case, the relevance of media practitioners becomes increasingly vital.

This study provides a pioneering examination of deepfake knowledge among media practitioners in the 2nd District of Cotabato Province, Philippines, a geographic and professional context that has not been previously studied in empirical research on synthetic media. Although public awareness of deepfakes has been the subject of other international research [8], [9], little attention has been paid to the particular knowledge gaps among frontline communicators in local media sectors in Southeast Asia.

Specifically, this study aimed to establish the socio-demographic characteristics of the respondents in terms of age, sex, civil status, number of years as media practitioners, and academic undergraduate program taken; determine respondents' knowledge on deepfake content, creation, detection, and software used for deepfake creation and detection. It also examined the correlation between years of experience or academic background and their level of knowledge on deepfakes. By examining these factors, the study provides information about how equipped local media professionals are to handle the growing threat of synthetic media and its implications on public trust and journalistic integrity.

Deepfakes hinder the verification of content and the maintenance of ethical norms in rapidly evolving news cycles [8]. When it is hard to distinguish between authentic and fraudulent media, misinformation is more likely to occur. This risk grows when content must be disseminated quickly. Ethical issues like privacy assaults, reputational harm, and the spread of misleading information make the situation worse. Deepfakes spread faster than fact-checks and retractions. This often compromises the need to correct falsehoods [10].

Accordingly, media organizations must adopt a holistic approach that integrates ethical responsibility, technological innovation, and ongoing training to address these challenges [11]. With the use of AI-driven detection techniques, the identification of manipulated information is enhanced. Sustained training in digital media literacy is essential for equipping journalists with the critical competencies required to interrogate digital content and mitigate emerging threats within increasingly complex information environments. Formulating clear ethical frameworks can support practitioners in validating information, protecting privacy, and rectifying inaccuracies quickly. Additionally, collaboration with fact-checkers and technology specialists can supplement verification procedures, while modifications to newsroom workflows can facilitate more in-depth content assessment.

Ultimately, this study offers valuable insights for various stakeholders. For policymakers, the findings can guide in the crafting of guidelines promoting responsible media practices and the ethical use of emerging technologies. For media organizations, the results serve as a basis for assessing workforce readiness and identifying areas that require improvement or intervention. For future researchers, the study provides a foundation for further empirical inquiry about deepfake awareness, detection, and media preparedness in the digital age.

2. RESEARCH METHOD

2.1. Research Design

This study employed a descriptive-correlational research design. The descriptive component presented and analyzed the respondents' socio-demographic characteristics. It also assessed their level of knowledge about deepfakes. The correlational component examined the relationship between selected socio-demographic variables, specifically, the number of years as a media practitioner and the undergraduate academic program taken, and the respondents' overall level of knowledge on deepfakes.

2.2. Population and Research Sample

This study was conducted among 25 media practitioners from 9 randomly selected radio stations in the second district of Cotabato Province. A cluster random sampling technique was employed, wherein 9 out of the total 15 radio stations in the district were selected as clusters. The respondents included news writers, DJs, reporters, and broadcasters, all considered media practitioners for the purpose of this study. Technicians were excluded, as the focus was on individuals directly involved in content creation and dissemination.

2.3. Research Instruments

The study employed a validated survey questionnaire as the primary data collection instrument. It obtained a Kuder-Richardson Formula 20 (KR-20) reliability coefficient of 0.813, reflecting a high level of internal consistency. The questionnaire was divided into two main sections. The first section gathered information on the respondents' socio-demographic characteristics, while the second part assessed their knowledge of deepfakes, organized into five categories: deepfake content, deepfake creation, deepfake detection, software used for deepfake creation, and software used for deepfake detection.

Each correctly answered item was assigned a score of one (1), while incorrect answers received a score of zero (0). Media practitioners who attained a score of 60% or more of the total possible points were described as having good knowledge, while those who scored below this level were reported as having poor knowledge.

To ensure content validity, the questionnaire underwent review by experts in the fields of communication and information technology. It was pilot tested on a small group of media practitioners outside the primary study population to assess item clarity, structure, and reliability. The instrument was enhanced based on the feedback from both the expert review and pilot testing.

2.4. Data Collection Techniques

Using a structured and validated questionnaire, data on the media practitioners' knowledge about deepfake content, creation, detection, and the software used for both creation and detection, were gathered. The instrument was personally administered to respondents from nine randomly selected radio stations in the second district of Cotabato Province. Respondents were informed about the study's objectives and obtained their informed consent to guarantee ethical adherence to the research process. To ensure a high response rate and preserve the integrity and completeness of the data, all completed questionnaires were immediately collected following administration.

2.5. Data Analysis Techniques

Descriptive statistics, such as frequency counts and percentage distributions, were used to present the respondents' socio-demographic characteristics and their level of knowledge about deepfakes. Spearman's rank correlation was utilized to analyze the relationship between years of experience as a media practitioner and knowledge of deepfakes, given the ordinal character of the variables. Moreover, Fisher's exact test was employed to assess the association between the respondents' undergraduate academic program and their level of knowledge about deepfakes. This method is appropriate considering the limited sample size and categorical data.

2.6. Research Procedures

Approval to conduct the study was first secured from the Research Adviser, Department Chairperson of Development Communication, Department Research Coordinator, and the University Research and Development Office (RDO). To seek permission to conduct an in-person survey with media practitioners, a formal request letter was handed to the management of the randomly selected radio station in Cotabato Province.

Prior to distributing the questionnaires, the study's objectives and methodology were clearly explained to the respondents. To ensure confidentiality and anonymity, each questionnaire included a Data Privacy Act notice. Additionally, informed consent was obtained from all willing respondents to uphold ethical standards during the research process.

3. RESULTS AND DISCUSSION

Findings of the research study is organized into three main sections: (1) the socio-demographic characteristics of the respondents, (2) their level of knowledge on deepfakes, and (3) the relationship between selected socio-demographic variables and the respondents' level of knowledge on deepfakes.

3.1. Socio-demographic Characteristics of the Respondents

The socio-demographic characteristics, such as age, sex, civil status, years of experience in the media industry, and undergraduate academic program taken by the 25 media practitioners included in the study, are presented in Table 1. Although the sample size is relatively small, it provides a significant overview of the media workforce in the 2nd District of Cotabato Province. It also serves as a basis for understanding how demographic variables may influence knowledge about deepfakes.

The respondents exhibited a relatively balanced age distribution across five groups: 24% were aged 20–25 years, 24% were aged 36–40 years, and 24% were aged 41–60 years. Moreover, 16% were within the age range of 26 to 30, while 12% were aged 31 to 35. This equitable representation indicates a varied generational viewpoint within the media workforce. Although younger media practitioners are often perceived to be more familiar with digital technologies [12], the data does not indicate a dominant age group, suggesting that both younger and older practitioners are actively engaged in media work.

Regarding sex, the sample showed a slight male predominance, with 64% of respondents identifying as male and 36% as female. In terms of civil status, the respondents were nearly evenly split, with 52% single and 48% married, suggesting that personal status may not be a major factor affecting engagement in the media profession.

On media professional experience, 40% of respondents indicated they have been in the industry for seven years or more, 32% for 5-6 years, 12% for 1-2 years, 8% for 3-4 years, and an additional 8% for less than a year. This signifies a robust blend of experienced professionals and recent arrivals, providing diverse viewpoints on the emerging difficulties posed by deepfake technologies. According to Chaudhuri, Groh, & Mehta [13], long-term exposure to media work may improve the awareness and detection of digital manipulations, including deepfakes; however, this is also affected by additional characteristics such as training and academic background.

Notably, 10 (40%) of the respondents finished non-communication undergraduate programs like Education (5), AB English (2), AB Philosophy (1), Bachelor of Technology (1), and BS Accountancy (1). The remaining 60% had formal academic training in communication-related fields: AB Mass Communication (28%), BS Development Communication (16%), AB Journalism (12%), and AB Communication (4%). This diversity in academic backgrounds reflects the different pathways into media practice and may affect practitioners' knowledge of digital media phenomena, such as deepfakes.

The diverse socio-demographic characteristics of media professionals in Cotabato Province highlight significant factors for tackling deepfake issues. Forty percent of the respondents lack formal communication education, indicating a notable gap in critical media literacy, which is essential for the detection of manipulated content [14]. Moreover, diverse experience levels indicate varying degrees of familiarity with developing technologies; seasoned professionals may depend on conventional procedures, but less experienced practitioners may require additional training in ethics and verification [15].

Demographic discrepancies meaningfully affect technology engagement and the perception of deepfakes [16], highlighting the necessity for customized capacity-building. Media organizations must establish extensive training programs that incorporate digital proficiencies and journalistic ethics, augmented by AI detection tools [17]. Cooperative initiatives among media, academia, and tech developers can increase resilience to misinformation [18].

From the data, it is implied that education and technological integration are crucial for increasing media practitioners' ability to combat deepfakes and sustaining public trust.

Table 1. Socio-demographic characteristics of the respondents. 2nd District, Cotabato Province, Philippines. 2024

Characteristics	Frequency (n=25)	Percentage (%)
Age		
20-25	6	24.00
26-30	4	16.00
31-35	3	12.00
36-40	6	24.00
41and up	6	24.00
Sex		
Male	16	64.00
Female	9	36.00
Civil Status		
Single	13	52.00
Married	12	48.00
Number of years as media practitioner		
7 years and above	10	40.00
5-6 years	8	32.00
3-4 years	2	8.00
1-2 years	3	12.00
Less than a year	2	8.00
Undergraduate academic program taken		

AB Mass Communication	7	28.00
BS Development Communication	4	16.00
AB Journalism	3	12.00
AB Comunication	1	4.00
Others (Education-related, AB English, Bachelor of Technology, AB Philosophy, and BS Accountancy	10	40.00

3.2. Deepfake Content

Table 2 presents the self-reported knowledge of 25 media practitioners regarding deepfake content, revealing both strengths and critical knowledge gaps. A strong majority (92%) correctly identified the potential for textual deepfakes to spread misinformation, aligning with Pawelec [19], who emphasizes that deepfakes threaten the credibility of scientific and factual communication by facilitating disinformation.

Further, high levels of understanding were evident in core deepfake concepts: 88% correctly defined textual and audio deepfakes, and 80% accurately recognized manipulated visuals and image-based deepfakes such as those generated by FaceApp and voice cloning as audio deepfakes. This reflects a foundational awareness of the various forms of deepfake media, which is essential for media professionals tasked with verifying content [8].

However, notable misconceptions emerged. A majority (88%) erroneously believed that deepfakes cannot be used for harmful purposes such as identity fraud or privacy invasion. This contradicts the statement of [20], who stressed that deepfakes have been exploited for malicious intents, including identity manipulation and reputational harm. Similarly, 88% of respondents failed to recognize the creative and benign applications of deepfakes, such as satire or entertainment, which Emmerson, Huang, & Taft [21] highlight as emerging uses in digital media.

Of further concern, 36% of respondents did not perceive deepfakes as a threat to journalistic integrity. This gap is significant, as Temir [22] and Kaate et al. [23] found that deepfakes can erode public trust and even appear more credible than genuine media, making journalists particularly vulnerable to inadvertently spreading false content. Additionally, 32% of respondents lacked awareness of the increasing sophistication of deepfake technology. Godulla, Hoffmann, & Seibert [5] warn that as deepfakes become more realistic, distinguishing manipulated content from authentic material will become increasingly difficult for journalists, putting editorial credibility at risk.

Overall, 64% of respondents were classified as having good knowledge, while 36% demonstrated poor understanding (Table 3). These results imply that while media practitioners in Cotabato Province have a baseline awareness of deepfake technologies, misconceptions remain, particularly regarding the dual-use nature of deepfakes (both harmful and creative), their evolving complexity, and the broader implications for media credibility. It suggests that media practitioners possess a foundational understanding of deepfakes; however, there is room for improvement regarding their awareness of the evolving nature of deepfakes, the potential for positive use cases, and the significant threat they pose to reliable information sources. These findings emphasize the need for ongoing education and training for media professionals to stay informed and adapt their practices in the face of this evolving technology.

The findings underscore the need for continuous professional development among media practitioners. Up-to-date knowledge on deepfake detection tools, ethical implications, and the broader socio-political impact of synthetic media needs to be integrated into designing training programs [15]. Doing so will help build media resilience, uphold journalistic standards, and equip practitioners to navigate an increasingly manipulated information environment.

Table 2. Knowledge on deepfake content. 2nd District, Cotabato Province, Philippines. 2024

Statements	Correct		Incorrect	
	Frequency (n=25)	Percentage (%)	Frequency (n=25)	Percentage (%)
1. I know that deepfake videos cannot be used for sarcastic and humorous purposes such as memes, and celebrity impersonation.	3	12	22	88
2. I know that textual deepfakes are texts that imitate human-like written content.	22	88	3	12
3. I know that deepfake audio is an audio that closely mimics a real person's voice.	22	88	3	12

4. I know that deepfake video is a fabricated video to mimic someone's physical attribution.	20	80	5	20
5. I know that textual deepfakes can be used to spread out misinformation.	23	92	2	8
6. I know that voice cloning is an example of audio deepfake.	20	80	5	20
7. I know that face app is an example of deepfake image.	20	80	5	20
8. I know that deepfake images cannot be used for detrimental purposes like forging false identities or violating someone's privacy.	3	12	22	88
9. I know that deepfakes undermine journalism and trustworthy sources of information.	16	64	9	36
10. I know that the quality of deepfakes is getting better over time.	17	68	8	32

Table 3. Level of knowledge on deepfake content. 2nd District, Cotabato Province, Philippines, 2024

Level of Knowledge	Frequency (n=25)	Percentage (%)
Good knowledge	16	64.00
Poor knowledge	9	36.00
Total	25	100.00

3.3. Deepfake Creation

This part presents the level of knowledge of media practitioners regarding the technical creation of deepfake content. The results indicate a moderate understanding among respondents, with 60% demonstrating good knowledge and 40% showing gaps in technical awareness (Table 5). Several key concepts were correctly identified, yet some fundamental misconceptions persist.

Most notably, 80% of respondents accurately understood that audio deepfakes rely on speech synthesis and recognition algorithms to mimic human voices (Statement 5), consistent with the findings of Kietzmann et al. [24], who noted the increasing use of voice cloning technologies in misinformation campaigns. Similarly, 72% correctly recognized that deepfakes can be generated by training AI models on large datasets of audio recordings to replicate a specific speaker's voice (Statement 3), which reflects the common use of machine learning techniques in speech mimicry [25].

Furthermore, 68% correctly acknowledged that real-time deepfakes can be created through video conferencing platforms and streaming technologies (Statement 2), a capability increasingly exploited for impersonation and fraud [26]. Respondents also demonstrated awareness of textual deepfakes, with 68% correctly identifying AI chatbots as tools for generating human-like text (Statement 8), aligning with growing concerns about large language models like ChatGPT being misused for disinformation [27].

However, significant misconceptions were also evident. A large majority (88%) incorrectly believed that facial recognition algorithms are not essential in deepfake video creation (Statement 7). This contradicts the work of Liu et al. [28], who emphasized that facial recognition and mapping algorithms, such as those used in FaceSwap and DeepFaceLab, are critical components of video-based deepfake generation using deep learning and computer vision technologies. This misunderstanding suggests the need for clearer training on how visual deepfakes are technically constructed.

Another common misconception (88%) was the belief that creating deepfakes requires advanced AI expertise (Statement 9). Masood et al. [29] clarified that with the growing availability of user-friendly applications and pre-trained models, the barrier to entry for creating deepfakes has significantly lowered, requiring minimal technical skill and financial investment. Interestingly, 76% of respondents correctly identified FaceApp as a tool capable of generating deepfake content (Statement 10), supporting [29]'s assertion that consumer-grade applications contribute to the rapid democratization of synthetic media creation.

Additionally, there was reasonable awareness of core AI techniques: 72% identified the creation of deepfake media through training machine learning models on large amounts of audio data to mimic a specific speaker (Statement 3), 68% of respondents recognized the three-step process in deepfake creation which are data collection, model training, and content generation (Statement 4), while 60% correctly identified the use of autoencoders and Generative Adversarial Networks (GANs) (Statement 6). Moreover, 56% understood the generator-discriminator model at the heart of GANs (Statement 1), reflecting an introductory grasp of how AI models refine synthetic outputs through iterative feedback loops [30].

Moreover, a significant portion of respondents (72%) correctly identified the creation of deepfake media through training machine learning models on large amounts of audio data to mimic a specific speaker (Statement 3), 68% agreed that there are three steps in creating deepfake material (data collection, training, and generation) (Statement 4), 60% of them recognized that the most common ways for deepfake creation are autoencoders and Generative Adversarial Networks (statement 6), and 56% recognized that deepfake material is generated by two competing algorithms, the generator-discriminator dynamic (Statement 1).

In summary, while a foundational understanding of deepfake creation processes exists among media practitioners, there remain substantial misconceptions regarding the technical requirements and accessibility of these tools. The findings underscore the importance of targeted media literacy training that not only highlights the social risks of deepfakes but also educates on their underlying mechanics. As Xiao & Huang [31] argue, equipping journalists and media professionals with technical knowledge is essential in combating the spread of synthetic content and maintaining public trust in media ecosystems.

Table 4. Knowledge on deepfake creation. 2nd District, Cotabato Province, Philippines. 2024

Statements	Correct		Incorrect	
	Frequency (n=25)	Percentage (%)	Frequency (n=25)	Percentage (%)
1. Deepfake material is generated by two competing AI algorithms, the generator and the discriminator.	14	56	11	44
2. A real-time or live deepfake can change one's face to another through video conferences and streaming networks.	17	68	8	32
3. Deepfake media are created by training machine learning models on large amounts of audio data to mimic a specific speaker.	18	72	7	28
4. Creating deepfake involves three steps: data collection, training, and generation.	17	68	8	32
5. Deepfake audio uses speech synthesis and speech recognition algorithms to imitate human voice or sound.	20	80	5	20
6. The most common way to generate deepfake is by using autoencoders and Generative Adversarial Networks.	15	60	10	40
7. Deepfake video does not need to use facial recognition algorithm to swap one person's face and replaces it with another.	3	12	22	88
8. Textual deepfake uses an artificial intelligence chatbot that can generate and understand natural human language.	17	68	8	32
9. Generating deepfake needs an advanced expertise in AI.	3	12	22	88
10. FaceApp utilizes neural networks to create deepfake images that simulate aging, gender swapping, hair alteration, and the way people smile.	19	76	6	24

Table 5. Level of knowledge on deepfake creation. 2nd District, Cotabato Province, Philippines. 2024

Level of Knowledge	Frequency (n=25)	Percentage (%)
Good knowledge	15	60.00
Poor knowledge	10	40.00
Total	25	100.00

3.4. Deepfake Detection

Table 2.5 illustrates the respondents' understanding of various methods used to detect deepfake content. The findings suggest that while there is partial awareness of general detection strategies, significant misconceptions remain, particularly regarding technical detection techniques, leading to an overall poor level of knowledge for 56% of the respondents (Table 7).

Notably, a large majority (92%) incorrectly believed that deepfakes cannot be detected through facial expression and blinking pattern analysis (Statement 1). This contradicts empirical research by Li et al. [32], who demonstrated that deepfake videos often fail to replicate natural blinking or consistent eye movement due to training data limitations. Similarly, 88% of respondents were unaware that deepfake audio can be detected by analyzing irregular speech patterns (Statement 8), despite findings by Almutairi & Elgibreen [33], which highlight that inconsistencies in prosody, pitch, and spectral features can expose audio manipulation.

A significant portion of respondents (72%) agreed that through identifying inconsistencies in face dimensions and body movements, deepfake materials can be detected (Statement 4). In the study conducted by Pan et al. [34] entitled "Deepfake Detection through Deep Learning", it was discussed that there are two main approaches to deepfake detection: forensic analysis and deep learning. Forensic analysis involves looking for inconsistencies in the video, such as blurring around the edges of the face or unrealistic skin tones.

On a more positive note, 72% of respondents correctly identified that inconsistencies in facial dimensions and unnatural body movements can serve as red flags in deepfake content (Statement 4). Heidari et al. [35] supported this through their work on forensic detection techniques, which emphasized artifacts such as blurred edges, mismatched lighting, and distorted facial geometry as indicators of deepfake generation.

Likewise, 72% correctly recognized that digital watermarking can aid in verifying media authenticity (Statement 5). Nadimpalli & Rattani [36] introduced a proactive watermarking strategy that involves embedding a visible watermark, generated through GANs, directly into original images. This ensures that if the image is later manipulated or transformed into a deepfake, the watermark remains detectable, allowing for clear identification of tampering.

Moreover, 68% understood the importance of metadata or content history in identifying manipulated media (Statement 3), echoing suggestions by Verdoliva [37] that provenance tracking can significantly strengthen media verification systems. A great portion of the respondents (60%) accurately recognized that deepfakes can be spotted by observing visual flaws like fake moles, odd lip color, unnatural skin texture, artificial facial hair, and shadow inconsistencies around the eyes (Statement 9), as to Amerini et al. [38], deepfake generators often fail to replicate fine facial details, which results in noticeable flaws such as uneven skin texture, unnatural lip pigmentation, missing or soft facial hair shadows, and lighting inconsistencies around the face. These subtle anomalies can serve as reliable indicators for identifying manipulated or synthetic media content.

Further, 56% acknowledged the use of deep learning models trained to identify subtle digital artifacts for detection (Statement 7), and a similar percentage agreed that platforms offering upload-and-detect functions can assist users in flagging fake videos (Statement 2). These insights align with the work of Afchar et al. [39], who demonstrated the efficiency of convolutional neural networks (CNNs) in differentiating real from synthetic faces.

However, significant knowledge gaps were still evident. A majority (56%) wrongly believed that deepfakes cannot be detected through pixel-based physiological signals such as blood flow (Statement 10), even though studies like Güera and Delp [40] have shown that minute color changes correlated with pulse can be used in deepfake identification. Another 56% dismissed the potential of detecting grayscale variations, subtle visual inconsistencies imperceptible to the naked eye (Statement 6), despite advancements in spectral analysis tools.

These findings suggest that while media practitioners are beginning to grasp basic detection strategies, they lack comprehensive knowledge of both forensic and AI-powered methods. This underscores the urgent need for robust media literacy training programs. As Verdoliva [37] emphasized, integrating education on detection strategies, both manual and automated, can significantly enhance public resilience against misinformation. Furthermore, Vizoso [41] noted that even advanced detection tools may lag behind as deepfake generation techniques become more sophisticated, highlighting an ongoing "technological arms race" between creators and detectors.

In conclusion, the generally poor level of knowledge (56%) among media practitioners regarding deepfake detection reveals a critical gap that must be addressed through capacity-building efforts. Empowering media personnel with updated knowledge on detection mechanisms is not just a professional necessity but a societal imperative to curb the proliferation of manipulated media in an increasingly digital information ecosystem.

Table 6. Knowledge on deepfake detection. 2nd District, Cotabato Province, Philippines. 2024

Statements	Correct		Incorrect	
	Frequency (n=25)	Percentage (%)	Frequency (n=25)	Percentage (%)
1. Deepfake cannot be detected through examining facial expressions and blinking patterns.	2	8	23	92

2. To spot deepfake material, users can simply provide a video link or upload the video for detection.	14	56	11	44
3. Maintaining a record of a piece of media's sources and history helps people detect if someone has tampered with it.	17	68	8	32
4. One way to detect deepfake is to identify inconsistencies in body and face dimensions, or between body movements or postures.	18	72	7	28
5. Adding a unique identifier to an image or text such as watermark makes it easier to track the source of media and it helps determine its authenticity.	18	72	7	28
6. A deepfake detection tool can detect subtle grayscale changes that are usually missed by normal eyes and provides real-time confidence scores for fast detection of deepfakes.	11	44	14	56
7. To detect deepfake videos, a deep learning model is usually trained to differentiate between real and fake information using patterns and artifacts seen in videos.	14	56	11	44
8. Deepfake audio cannot be detected by identifying irregularities in speech patterns.	3	12	22	88
9. Deepfake may be identified by paying close attention to details, such fictitious moles, unusual lip color, excessively smooth or wrinkled skin, artificial facial hair, and dull shadows around the eyes.	15	60	10	40
10. Real-time deepfake detector determines the authenticity of a video by analyzing the "blood flow" in the pixels.	11	44	14	56

Table 7. Level of knowledge on deepfake detection. 2nd District, Cotabato Province, Philippines. 2024.

Level of Knowledge	Frequency (n=25)	Percentage (%)
Poor knowledge	14	56.00
Good knowledge	11	44.00
Total	25	100.00

3.5. Software Used for Deepfake Creation

Table 2.7 presents the respondents' awareness of various software tools used in creating deepfake content. The findings underscore a relatively good level of knowledge among media practitioners, with 80% demonstrating adequate awareness of deepfake creation tools, while 20% exhibited poor knowledge (Table 9). This overall awareness reflects the increasing exposure of media practitioners to digital content manipulation technologies, though important gaps remain.

Specifically, a significant majority (88%) correctly identified DeepFaceLive as a tool used to generate real-time or live deepfake videos, indicating strong familiarity with facial reenactment software. Similarly, most respondents recognized FaceApp (72%) as an image manipulation tool and FaceSwap (68%) as software for video-based face-swapping, two applications that have gained mainstream popularity and have been featured in recent studies on deepfake accessibility [12], [37].

Awareness of Lyrebird AI, a tool for voice cloning and synthetic audio creation, was correctly identified by 52% of the respondents, indicating moderate familiarity with audio deepfake technologies. Similarly, 52% correctly recognized that ChatGPT, while not originally designed for malicious use, can be utilized for generating persuasive fake text or impersonating individuals in written form, contributing to so-called *textual deepfakes* [42], [43]. However, nearly half (48%) of the respondents failed to identify Lyrebird AI and ChatGPT as deepfake-related tools, highlighting a knowledge gap in non-visual manipulation technologies.

Specifically, a significant portion of respondents (88%) correctly identified that DeepFaceLive is used to create real-time or live deepfakes. The majority recognized that FaceApp (72%) is used for deepfake images and FaceSwap (68%) for deepfake videos. Awareness of Lyrebird AI for deepfake audio creation was correctly identified by 13 (52%) respondents; the same with ChatGPT (52%), which is used for textual deepfakes. However, almost half of the population of the respondents (48%) did not recognize Lyrebird AI and ChatGPT as software used for deepfake creation.

These findings reveal that while media practitioners show strong awareness of popular video and image-based deepfake tools, their understanding of audio and text-based software remains limited. This gap is significant, as deepfakes are no longer restricted to visual formats. Synthetic audio and AI-generated texts are increasingly used to spread misinformation in political and social contexts [9].

Although 80% of respondents demonstrated good overall knowledge, ongoing education is essential. As deepfake technologies rapidly evolve, new tools will emerge that may bypass current detection methods, requiring media professionals to stay informed and adaptive [26]. From a communication and policy perspective, enhancing media literacy, particularly in emerging deepfake forms, is crucial. Journalists and practitioners play a key role in identifying manipulated content and maintaining public trust [44].

While awareness of video/image-based deepfake tools is high, limited knowledge of audio and textual deepfakes highlights the need for continuous, interdisciplinary media literacy efforts. Targeted public education and institutional training are essential to keep pace with AI-driven misinformation and protect the integrity of information ecosystems.

Table 8 Knowledge on deepfake detection. 2nd District, Cotabato Province, Philippines. 2024

Statements	Correct		Incorrect	
	Frequency (n=25)	Percentage (%)	Frequency (n=25)	Percentage (%)
1. I know that FaceApp is used to create deepfake image.	18	72	7	28
2. I know that ChatGPT is used to create textual deepfake.	13	52	12	48
3. I know that Lyrebird AI is used to create deepfake audio.	13	52	12	48
4. I know that FaceSwap is used to create deepfake video.	17	68	8	32
5. I know that DeepFaceLive is used to create real-time or live deepfake.	22	88	3	12

Table 9. Level of knowledge on deepfake detection. 2nd District, Cotabato Province, Philippines. 2024.

Level of Knowledge	Frequency (n=25)	Percentage (%)
Good knowledge	20	80.00
Poor knowledge	5	20.00
Total	25	100.00

3.6. Software Used for Deepfake Detection

Table 10 shows that while some respondents were familiar with prominent deepfake detection tools, overall awareness remains limited. Specifically, 64% correctly identified WeVerify as a tool for detecting manipulated images. However, a considerable portion of respondents misidentified tools like Sensity and Sentinel AI, with only 44% correctly associating them with textual and audio deepfakes, respectively. Similarly, tools such as Intel's FakeCatcher and Deepware Scanner, both used in video deepfake detection, were correctly identified by just 56% of respondents.

These results indicate that although media practitioners have some knowledge of mainstream detection software, their understanding of specialized tools and their functionalities is inconsistent. This lack of clarity highlights a critical gap in digital literacy. As Haimson [45] emphasized, recognizing the distinct purposes of detection tools, whether for video, image, audio, or text, is essential for navigating today's complex media environment.

The generally poor knowledge level (52%) among respondents, as shown in Table 2.10, signals an urgent need for targeted awareness campaigns and institutional training. As deepfake technologies evolve, so do the tools required to detect them. According to Verdoliva [37], the growing sophistication of deepfakes necessitates a parallel advancement in detection literacy to prevent the spread of misinformation. Without a solid grasp of detection tools like FakeCatcher, Phoneme-Viseme Mismatch Detectors, or Microsoft Video AI Authenticator, media practitioners risk missing subtle manipulations that undermine journalistic integrity and public trust.

The findings underscore the necessity of integrating deepfake detection tools into media education curricula. Equipping media professionals with up-to-date, tool-specific knowledge will enhance their capacity to verify digital content and counteract the spread of synthetic media. In line with Maras & Alexandrou [15], digital literacy must evolve beyond general awareness to include technical understanding of detection technologies as part of a broader strategy to combat disinformation.

Table 10. Knowledge on deepfake detection. 2nd District, Cotabato Province, Philippines. 2024

Statements	Correct		Incorrect	
	Frequency (n=25)	Percentage (%)	Frequency (n=25)	Percentage (%)
1. I know that Intel's FakeCatcher is used to detect real-time deepfake.	14	56	11	44
2. I know that WeVerify is used to detect deepfake image.	16	64	9	36
3. I know that Deepware Scanner is used to detect deepfake videos.	14	56	11	44
4. I know that Sensity is used to detect textual deepfakes.	11	44	14	56
5. I know that Sentinel AI is used to detect deepfake audio.	11	44	14	56

Table 11. Level of knowledge on deepfake detection. 2nd District, Cotabato Province, Philippines. 2024.

Level of Knowledge	Frequency (n=25)	Percentage (%)
Poor knowledge	13	52.00
Good knowledge	12	48.00
Total	25	100.00

3.7. Relationship between Respondents' Socio-demographic Characteristics and Level of Knowledge on Deepfakes

Table 4 presents the correlation analysis between selected socio-demographic characteristics and the respondents' level of knowledge on deepfakes. Results reveal a significant positive correlation between the number of years as a media practitioner and knowledge on deepfake content ($p = .02 < .05$), with a moderate correlation coefficient ($rs = .474$). This suggests that experience in the media field enhances the practitioner's ability to understand deepfake content. These findings correspond with Blancaflor et al. [46], who highlighted that individuals with greater exposure to media and digital manipulation possess enhanced capabilities to identify and evaluate altered content, owing to their refined analytical and visual literacy skills.

On the other hand, no significant correlations were identified between the number of years in media practice and knowledge related to deepfake creation ($p = .42$), detection ($p = .43$), software used for creation ($p = .29$), and software used for detection ($p = .09$). This implies that although expertise may facilitate content recognition, it does not inherently correspond to proficiency with the technical elements or instruments related to deepfake creation and identification. This highlights the necessity for specialized training that encompasses both practical experience and technical skills.

As regards to academic background, the analysis revealed no significant correlations between the respondents' undergraduate programs taken and any dimension of deepfake knowledge including content ($p = 1.00$), creation ($p = .442$), detection ($p = 1.00$), software used for creation ($p = .615$), and software used for detection ($p = 1.00$). These findings suggest that formal education, regardless of discipline, may not significantly influence one's awareness or understanding of deepfakes. This supports the findings of Birrer and Just [47], who reported that attributes such as age, trust in information sources, and media literacy are more significant predictors of vulnerability than whether someone studied computer science vs. non-technical fields. Their systematic review found no consistent patterns linking formal education or academic discipline to improved ability to detect deepfakes.

These results highlight the value of practical media experience in developing deepfake awareness, particularly in recognizing manipulated content. However, the lack of correlation between experience or educational background and technical knowledge (e.g., detection tools and software) indicates a pressing need for structured, interdisciplinary media literacy programs. As emphasized by Mirsky and Lee [26], the evolving nature of synthetic media necessitates continuous upskilling of media professionals, regardless of academic or experiential background, to effectively detect and mitigate the risks associated with AI-generated disinformation.

Table 4. Test of relationship between socio-demographic characteristics and level of knowledge on deepfakes. 2nd District, Cotabato Province, Philippines. 2024

Variable		Coefficient	Interpretation	P-value	Conclusion
Number of years as media practitioner	Knowledge				
	Content	.474*	Positive moderate	0.02	Significant
	Creation	0.17^{ns}	Positive weak	0.42	Not significant
	Detection	0.17^{ns}	Positive weak	0.43	Not significant
	Software used for Deepfake Creation	0.22^{ns}	Positive weak	0.29	Not significant
	Software used for Deepfake Detection	0.35^{ns}	Positive weak	0.09	Not significant
Undergraduate academic program taken	Knowledge				
	Content			1.000	Not significant
	Creation			0.442	Not significant
	Detection			1.000	Not significant
	Software used for Deepfake Creation			0.615	Not significant
	Software used for Deepfake Detection			1.000	Not significant

3.8. Implications and Limitations of The Study

The findings of this study suggest that while media practitioners in Cotabato Province possess a foundational understanding of deepfake content, particularly in its form and basic characteristics, they lack sufficient technical knowledge in areas of creation, detection, and software tools. The significant positive correlation between years of media experience and content knowledge reinforces the value of experiential learning in navigating media manipulation [46]. However, the absence of correlations with technical knowledge domains highlights a critical gap that cannot be bridged by experience alone and calls for targeted digital literacy interventions.

This study reinforces earlier research emphasizing the challenges deepfakes pose to journalistic integrity, public trust, and media verification workflows [8], [22], [23]. Despite high awareness of visual deepfakes, media practitioners showed limited familiarity with text and audio-based manipulations, echoing concerns about the rising misuse of AI-generated text and synthetic voices for misinformation [43], [27]. Further, poor understanding of deepfake detection methods and software tools suggests a need for institutional training programs, aligning with calls by Verdoliva [37] and Haimson [45] for strengthening detection literacy.

Limitations of the study include the relatively small sample size (n = 25), which may limit the generalizability of the findings to broader populations of media practitioners. Additionally, the study relied on self-reported data, which may introduce biases related to over - or underestimation of deepfake knowledge. Finally, the study focused on one geographic district, which may not reflect the diversity of media literacy across other regions in the Philippines. Future studies could explore a comparative analysis across multiple provinces, integrate observational or task-based assessments of deepfake detection skills, and investigate the effectiveness of specific training interventions in improving deepfake resilience among journalists.

4. CONCLUSION

This study found that media practitioners in District 2, Cotabato Province, Philippines, are generally aware of the existence of deepfakes and demonstrate a degree of familiarity with the tools used in their creation. However, significant gaps remain in their ability to detect deepfakes, particularly in recognizing detection software and distinguishing manipulative content from authentic media. Statistical analysis revealed a significant relationship between years of media work experience and knowledge of deepfake content, indicating that professional experience can increase awareness of the issue. Therefore, the null hypothesis was rejected for this variable. However, no significant relationship was found between years of experience and knowledge of deepfake creation, detection, or software. Similarly, respondents' undergraduate educational background did not significantly influence deepfake literacy among media practitioners.

These findings underscore the importance of experiential learning in the media workplace and the need for more targeted education and capacity building to address emerging digital threats like deepfakes. Based on the results of this study, several recommendations can be put forward, including: (1) policymakers need to design and implement regulations related to the ethical use of AI and digital technology, particularly in preventing misinformation and regulating deepfake practices; (2) media organizations can organize regular capacity-building activities such as seminars, training, and simulations on deepfake detection, as well as encourage knowledge transfer between generations of practitioners; (3) further researchers can examine gender variables in understanding perceptions and competencies related to deepfake threats, considering that the majority of respondents to this study were male (64%); and (4) replicate the study across a wider geographic scope, for example across Cotabato Province, to gain more comprehensive insights into the preparedness of media practitioners in dealing with the deepfake phenomenon.

REFERENCES

[1] A. Aggarwal and R. Garg, "Fake news on social media: The impact on society," *Information Systems Frontiers*, vol. 24, pp. 1191–1210, 2022. [Online]. Available: https://doi.org/10.1007/s10796-022-10242-z
[2] N. Paunović, "Disclosing and disseminating fake news through media and freedom of expression," in *Mediji, Kazneno Pravo i Pravosuđe*, Belgrade, Serbia: Institute of Criminological and Sociological Research, 2024, pp. 107–116. [Online]. Available: https://www.iksi.ac.rs/izdanja/mediji_kazneno_pravo_i_pravosudje_2024.pdf#page=107
[3] World Economic Forum, *Global Risks Report 2022: Emerging Technologies and Security Threats*, Jan. 2022. [Online]. Available: https://www.weforum.org/publications/global-risks-report-2022/
[4] X.-J. Lim, J. Zhu, D. Leong, and S. Foo, "Deepfake awareness and public perception: The gap between concern and detection ability," *Cyberpsychology, Behavior, and Social Networking*, vol. 24, no. 3, pp. 170–175, 2021. [Online]. Available: https://doi.org/10.1089/cyber.2020.0100
[5] A. Godulla, C. P. Hoffmann, and D. Seibert, "Dealing with deepfakes–an interdisciplinary examination of the state of research and implications for communication studies," *SCM Studies in Communication and Media*, vol. 10, no. 1, pp. 72–96, 2021. [Online]. Available: https://doi.org/10.5771/2192-4007-2021-1-72
[6] K. Dunwoodie, L. Macaulay, and A. Newman, "Qualitative interviewing in the field of work and organisational psychology: Benefits, challenges and guidelines for researchers and reviewers," *Applied Psychology*, vol. 72, no. 2, pp. 863–889, 2023. [Online]. Available: https://doi.org/10.1111/apps.12414
[7] D. Fallis, "The epistemic threat of deepfakes," *Philosophy & Technology*, vol. 34, no. 4, pp. 623–643, 2021. [Online]. Available: https://doi.org/10.1007/s13347-020-00419-2
[8] C. Vaccari and A. Chadwick, "Deepfakes and disinformation: Exploring the impact of synthetic political video on uncertainty and trust in news," *Social Media + Society*, vol. 6, no. 1, 2020. [Online]. Available: https://doi.org/10.1177/2056305120903408
[9] M. Westerlund, "The emergence of deepfake technology: A review," *Technology Innovation Management Review*, vol. 9, no. 11, pp. 40–53, 2019. [Online]. Available: https://doi.org/10.22215/timreview/1282
[10] D. M. West, "How to combat fake news and disinformation," *Brookings Institution*, 2021. [Online]. Available: https://www.brookings.edu/research/how-to-combat-fake-news-and-disinformation/
[11] B. Paris and J. Donovan, "Deepfakes and cheap fakes: The manipulation of audio and visual evidence," *Data & Society*, 2019. [Online]. Available: https://datasociety.net/library/deepfakes-and-cheap-fakes/
[12] T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi, "Deep learning for deepfakes creation and detection: A survey," *arXiv preprint arXiv:1909.11573*, 2019. [Online]. Available: https://arxiv.org/abs/1909.11573
[13] A. Chaudhuri, M. Groh, and R. Mehta, "Human performance in detecting deepfakes: A systematic review and meta-analysis of 56 papers," *Computers in Human Behavior Reports*, vol. 16, p. 100538, 2024. [Online]. Available: https://doi.org/10.1016/j.chbr.2024.100538
[14] E. C. Tandoc, Z. W. Lim, and R. Ling, "Defining 'Fake News': A typology of scholarly definitions," *Digital Journalism*, vol. 6, no. 2, pp. 137–153, 2018. [Online]. Available: https://doi.org/10.1080/21670811.2017.1360143
[15] M. H. Maras and A. Alexandrou, "Determining authenticity of video evidence in the age of artificial intelligence and in the wake of deepfake videos," *The International Journal of Evidence & Proof*, vol. 23, no. 3, pp. 255–262, 2019. [Online]. Available: https://doi.org/10.1177/1365712718807226
[16] V. Frissen, M. Van Lieshout, and L. Kool, "Deepfakes: Tackling threats and strengthening democracy," *Journal of Information Technology & Politics*, vol. 18, no. 4, pp. 355–369, 2021. [Online]. Available: https://doi.org/10.1080/19331681.2021.1950705
[17] S. Johnson and J. Seaton, "Journalism and misinformation in the age of deepfakes," *Digital Journalism*, vol. 9, no. 3, pp. 378–394, 2021. [Online]. Available: https://doi.org/10.1080/21670811.2020.1848955
[18] D. M. West, J. R. Allen, and J. Gibbs, "How artificial intelligence is transforming the world," *Brookings Institution*, 2019. [Online]. Available: https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/
[19] M. Pawelec, "Deepfakes and democracy (theory): How synthetic audio visual media for disinformation and hate speech threaten core democratic functions," *Digital Society*, vol. 1, no. 2, p. 19, 2022. [Online]. Available: https://doi.org/10.1007/s44206-022-00010-6
[20] T. Trinh and J. Liu, "Manipulating faces for identity theft via morphing and deepfake: Digital privacy," in *Handbook of Statistics*, vol. 48, pp. 223–241, 2023. [Online]. Available: https://doi.org/10.1016/bs.host.2022.12.003
[21] K. Emmerson, Y. Huang, and J. Taft, "Creative media applications of synthetic voices and deepfakes: A typology and research agenda," *AI & Society*, vol. 38, pp. 107–118, 2023. [Online]. Available: https://doi.org/10.1007/s00146-024-02072-1
[22] E. Temir, "Deepfake: New Era in The Age of Disinformation & End of Reliable Journalism," *Selçuk İletişim*, vol. 13, no. 2, pp. 1009–1024, 2020. [Online]. Available: https://doi.org/10.18094/JOSC.685338
[23] I. Kaate, J. Salminen, S. G. Jung, H. Almerekhi, and B. J. Jansen, "How do users perceive deepfake personas? Investigating deepfake user perception and its implications for HCI," in *Proc. ACM Int. Conf.*, 2023. [Online]. Available: https://doi.org/10.1145/3605390.3605397
[24] J. H. Kietzmann, I. P. McCarthy, L. W. Lee, and H. Bugshan, "Deepfakes: Trick or treat?," *Business Horizons*, vol. 63, no. 2, pp. 135–146, 2020. [Online]. Available: https://doi.org/10.1016/j.bushor.2019.11.006

[25] J.-E. Choi, K. Schäfer, and S. Zmudzinski, “Introduction to audio deepfake generation: Academic insights for non experts,” in *Proc. 3rd ACM Int. Workshop Multimedia AI Against Disinformation (MAD ’24)*, pp. 1–10, 2024. [Online]. Available: https://doi.org/10.1145/3643491.3660286

[26] Y. Mirsky and W. Lee, “The creation and detection of deepfakes: A survey,” *ACM Comput. Surv.*, vol. 54, no. 1, pp. 1–41, 2021. [Online]. Available: https://doi.org/10.1145/3425780

[27] L. Weidinger, J. Mellor, M. Rauh, C. Griffin, J. Uesato, P.-S. Huang, M. Cheng, M. Glaese, B. Balle, A. Kasirzadeh, Z. Kenton, S. Brown, W. Hawkins, T. Stepleton, C. Biles, A. Birhane, J. Haas, R. Rimell, L. A. Hendricks, and I. Gabriel, “Ethical and social risks of harm from language models: A framework and classification,” *arXiv preprint* arXiv:2112.04359, 2021. [Online]. Available: https://arxiv.org/abs/2112.04359

[28] K. Liu, I. Perov, D. Gao, N. Chervoniy, K. Liu, and W. Zhang, “DeepFaceLab: Integrated, flexible and extensible face-swapping framework,” *Pattern Recognit.*, vol. 141, Art. no. 109628, 2023. [Online]. Available: https://doi.org/10.1016/j.patcog.2023.109628

[29] M. Masood, M. Nawaz, K. M. Malik, A. Javed, A. Irtaza, and H. M. Malik, “Deepfakes generation and detection: State of the art, open challenges, countermeasures, and way forward,” *Appl. Intell.*, vol. 52, pp. 11233–11263, 2022. [Online]. Available: https://doi.org/10.1007/s10489-022-03766-z

[30] I. J. Goodfellow *et al.*, “Generative adversarial nets,” in *Advances in Neural Information Processing Systems*, vol. 27, Z. Ghahramani *et al.*, Eds. Curran Associates, 2014, pp. 2672–2680. [Online]. Available: https://doi.org/10.5555/2969033.2969125

[31] A. Xiao and X. Huang, “Wired to seek, comment and share? Examining the relationship between personality, news consumption and misinformation engagement,” *Online Information Review*, vol. 46, no. 9, pp. XX–XX, 2022. [Online]. Available: https://doi.org/10.1108/OIR-10-2021-0520

[32] Y. Li, M.-C. Chang, and S. Lyu, “In Ictu Oculi: Exposing AI generated fake face videos by detecting eye blinking,” in *Proc. 2018 IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, 2018, pp. 1–7. [Online]. Available: https://doi.org/10.1109/WIFS.2018.8630787

[33] Z. Almutairi and H. Elgibreen, “A review of modern audio deepfake detection methods: Challenges and future directions,” *Algorithms*, vol. 15, no. 5, p. 155, 2022. [Online]. Available: https://doi.org/10.3390/a15050155

[34] D. Pan, L. Sun, R. Wang, X. Zhang, and R. O. Sinnott, “Deepfake Detection through Deep Learning.” *In Proceedings of the 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT)*, Leicester, UK, Dec. 7–10, 2020. DOI: https://doi.org/10.1109/BDCAT50828.2020.00001

[35] A. Heidari, N. Jafari Navimipour, H. Dag, and M. Unal, “Deepfake detection using deep learning methods: A systematic and comprehensive review,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 14, no. 2, p. e1520, 2024. [Online]. Available: https://doi.org/10.1002/widm.1520

[36] A. V. Nadimpalli and A. Rattani, “ProActive DeepFake Detection using GAN-based Visible Watermarking,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 20, no. 11, Article 344, pp. 1–27, Sept. 2024. [Online]. Available: https://doi.org/10.1145/3625547

[37] L. Verdoliva, “Media forensics and deepfakes: An overview,” *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 5, pp. 910–932, 2020. [Online]. Available: https://doi.org/10.1109/JSTSP.2020.3002101

[38] I. Amerini, M. Barni, S. Battiato, P. Bestagini, G. Boato, T. S. Bonaventura, V. Bruni, R. Caldelli, F. De Natale, R. De Nicola, L. Guarnera, S. Mandelli, G. L. Marcialis, M. Micheletto, A. Montibeller, G. Orrù, A. Ortis, P. Perazzo, G. Puglisi, D. Salvi, *et al.*, “Deepfake media forensics: Status and future challenges,” *Journal of Imaging*, vol. 11, no. 3, Art. no. 73, 2025. [Online]. Available: https://doi.org/10.3390/jimaging11030073

[39] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, “MesoNet: A compact facial video forgery detection network,” in *Proc. 2018 IEEE Int. Workshop Inf. Forensics Security (WIFS)*, 2018, pp. 1–7. [Online]. Available: https://doi.org/10.1109/WIFS.2018.8630761

[40] D. Guera and E. J. Delp, “Deepfake video detection using recurrent neural networks,” in *Proc. 2018 IEEE Int. Conf. Adv. Video Signal Based Surveillance (AVSS)*, 2018, pp. 1–6. [Online]. Available: https://doi.org/10.1109/AVSS.2018.8639163

[41] A. Vizoso, M. Vaz Álvarez, and X. López García, “Fighting deepfakes: Media and internet giants’ converging and diverging strategies against hi tech misinformation,” *Media and Communication*, vol. 9, no. 1, pp. 291–300, 2021. [Online]. Available: https://doi.org/10.17645/mac.v9i1.3494

[42] L. Floridi and M. Chiriatti, “GPT-3: Its nature, scope, limits, and consequences,” *Minds and Machines*, vol. 30, no. 4, pp. 681–694, 2020. [Online]. Available: https://doi.org/10.1007/s11023-020-09548-1

[43] S. Kreps, R. McCain, and M. Brundage, “All the news that’s fit to fabricate: AI-generated text as a tool of media misinformation,” *Journal of Experimental Political Science*, vol. 9, no. 1, pp. 104–117, 2022. [Online]. Available: https://doi.org/10.1017/XPS.2021.28

[44] E. C. Tandoc, Z. W. Lim, and R. Ling, “Defining ‘fake news’: A typology of scholarly definitions,” *Digital Journalism*, vol. 6, no. 2, pp. 137–153, 2018. [Online]. Available: https://doi.org/10.1080/21670811.2017.1360143

[45] O. L. Haimson, “Deepfake detection and the evolving digital threat landscape: Implications for media professionals,” *Journal of Media Literacy Education*, vol. 15, no. 1, pp. 45–58, 2023. [Online]. Available: https://doi.org/10.23860/JMLE-2023-15-1-4

[46] E. Blancaflor, A. P. Ong, A. L. E. Navarro, K. F. Sudo, D. A. Villasor, and C. Valero, “An in-depth analysis of revenge porn and blackmailing on Philippine social media and its effects on the people affected,” in *Proc. 2023 6th Int. Conf. Inf. Sci. Syst.* (ICISS), 2023, pp. 107–112. [Online]. Available: https://doi.org/10.1145/3625156.3625172

[47] A. Birrer and N. Just, “What we know and don’t know about deepfakes: An investigation into the state of the research and regulatory landscape,” *New Media & Society*, 2024. [Online]. Available: https://doi.org/10.1177/14614448241253138