SYNTHETIC MEDIA AS A THREAT TO CRISIS COMMUNICATION: AI-GENERATED VIDEOS, DISINFORMATION, AND ALGORITHMIC AMPLIFICATION IN EMERGENCY SITUATIONS

Ljilja Šikman[1], Isidora Kojić[2], Biljana Vranješ[3], Đorđe Stojisavljević[4], Aleksandar Gaćina[4]

[1]University of Banjaluka, Faculty of Technology, Bulevar vojvode Stepe Stepanovića 73, 78000Banjaluka, Bosnia and Herzegovina, ljilja.sikman@tf.unibl.org
[2]University of Banjaluka, Faculty of security science, Bulevar vojvode Živojina Mišića 10A, 78000Banjaluka, Bosnia and Herzegovina, isidora.kojic@fbn.unibl.org
[3]University of Banjaluka, Faculty of Mechanical Engineering, Bulevar vojvode Stepe Stepanovića 71,78000 Banjaluka, Bosnia and Herzegovina, biljana.vranjes@mf.unibl.org
[4]University of Banjaluka, University Computer Center, Bulevar vojvode Petra Bojovića 1A, 78000Banjaluka Bosnia and Herzegovina, djordje.stojisavljevic@unibl.org, aleksandar.gacina@unibl.org

ORIGINAL SCIENTIFIC PAPER

Received: 04 November 2026.
Accepted: 14 November 2026.
Published: 16 November 2026.
DOI: https://doi.org/10.63395/STEDConf15022026180S339

Corresponding Author: Ljilja Šikman, University of Banjaluka, Faculty of Technology, Bulevar vojvode Stepe Stepanovića 73, 78000Banjaluka, Bosnia and Herzegovina, ljilja.sikman@tf.unibl.org

Copyright © 2026 Ljilja Šikman et al.; published by UNIVERSITY PIM. This work licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.

Citation: Šikman, LJ., et al. (2026). Synthetic media as a threat to crisis communication: AI-generated videos, disinformation, and algorithmic amplification in emergency situations. In Kojić, D. & Sredić. D. (eds.) *Conference proceedings, XV International Conference on Social and Technological Development: STED 2026* (pp. 339-349). Trebinje, B&H: University PIM.

ABSTRACT

The development of generative artificial intelligence and its multimodal capabilities—particularly in the domain of producing realistic video content and synthesized speech—constitutes one of the most significant contemporary threats to effective crisis communication. This paper examines the ways in which AI-generated video and audio content function as vectors of disinformation in the context of natural disasters and crisis situations, and analyzes the mechanisms of their algorithmic dissemination across platforms such as TikTok, YouTube, Instagram, and X (Twitter). Adopting a theoretical-analytical approach, and combining a review of relevant academic literature, reports by international organizations, and case studies—including Hurricane Melisa and the associated AI-generated viral content—the paper maps the key risks this phenomenon poses to crisis management. Particular attention is devoted to the psychological mechanisms that render AI-generated visual and auditory content especially persuasive under conditions of heightened anxiety, as well as to institutional and technological responses to these threats. These include synthetic media detection tools, regulatory frameworks such as the EU AI

Act and the Digital Services Act, and the C2PA initiative for digital media authenticity verification. The paper concludes that there is no single technical solution capable of effectively suppressing the spread of AI-generated disinformation in real time, and that a comprehensive approach is required—one that integrates technological tools, regulatory measures, proactive institutional communication, and investment in public digital literacy. **Keywords:** generative artificial intelligence, deepfake, crisis communication, disinformation, crisis management

INTRODUCTION

The development of generative artificial intelligence—particularly tools for creating realistic video content such as OpenAI's Sora system and Google's Gemini—has enabled the creation of convincing synthetic content within seconds, without the need for specialized knowledge. This accessibility poses a serious challenge to crisis communication, whose effectiveness depends on the timely and accurate flow of information. A seminal study published in the journal *Science* demonstrated that false information on social media spreads farther, faster, and more broadly than truthful information, particularly when it elicits strong emotional reactions [1] — a pattern that AI-generated content further amplifies, contributing to the phenomenon that the RAND Corporation refers to as "truth decay" [2].

Natural disasters have proven to be particularly fertile ground for this type of disinformation due to the information vacuum preceding official announcements and the algorithmic prioritization of high-engagement content [3]. An analysis conducted by Logically Intelligence, which examined eight major disasters in the United States during the 2024–2025 period, confirms that AI-generated content systematically accelerates the spread of disinformation [4]. This is illustrated by Hurricane Melissa (2025), during which AI-generated videos of sharks in flooded swimming pools accumulated millions of views within a matter of hours [5], as well as the Los Angeles wildfires, where fabricated footage of the Hollywood Sign engulfed in flames received more than 12.6 million views before fact-checking organizations confirmed its AI-generated nature [6]. The aim of this paper is, through a theoretical-analytical approach and a combination of literature review and case studies, to map the mechanisms through which AI-generated content functions as a vector of disinformation during crisis situations and to analyze the existing technological, regulatory, and institutional responses to this threat.

KEY FIGURES: AI-GENERATED DISINFORMATION

The scale of the problem of AI-generated disinformation during natural disasters is best illustrated by quantitative indicators collected during recent crisis events in 2024 and 2025 (Table 1) [4], [5], [6], [7].

Table 1. Key indicators of the scale of AI-generated disinformation during selected natural disasters

Key data	Scope of the problem
Prereviews of AI -generated video content - Hurricane Melissa	3.6M+
Prereviws of AI -generated video content – Los Angeles fires	12.6M
Sharing AI shark videos – Instagram	200K+
Number of hours to reach viral spread	<24h
Disaster monitoring 2024 – 2025 (Logically AI)	8

The data indicate that AI-generated content during natural disasters does not represent an isolated or marginal phenomenon but rather a systematic pattern that recurs from one crisis to another. During Hurricane Melissa, videos depicting non-existent damage in Jamaica —including the well-known footage of sharks in a flooded hotel swimming pool—accumulated more than 3.6 million views within a matter of hours, while the shark footage itself was further disseminated

through thousands of individual shares on Instagram [8]. Similarly, during the Los Angeles wildfires in January 2025, fabricated footage of the Hollywood Sign engulfed in flames and other AI-generated content related to the wildfires reached a total of more than 12.6 million views across Facebook, TikTok, X, and Instagram before independent fact-checking organizations confirmed their artificial nature [7]; [6].

Of particular significance is the finding that AI-generated content typically reaches viral dissemination within less than 24 hours of the onset of a disaster—a timeframe shorter than the standard verification and publication cycle of official information by the competent authorities. This temporal asymmetry, confirmed by an analysis conducted by Logically Intelligence covering a total of eight major natural disasters during the 2024–2025 period, represents one of the central structural challenges addressed in this paper [9].

CASE STUDY 1: FAKE HURRICANE VIDEOS FEATURING AI-GENERATED SHARKS

Hurricane Melissa, which struck the Caribbean in late October 2025—with particularly severe consequences for Jamaica—represents a turning point in the phenomenon of AI-driven disinformation during natural disasters. It was the first major natural disaster to occur following the public release of the latest version of OpenAI's Sora video generation system, enabling researchers to observe, for the first time under real-world conditions, the scale and speed at which high-quality AI-generated content spreads during a crisis situation [5].

As the storm moved through the region, two particularly viral videos rapidly spread across social media platforms, primarily TikTok ((Figure 1), X, and Instagram [22].



Figure 1. AI generated video content on TikTok during Hurricane Melissain in 2025

The first depicted four sharks swimming through the swimming pool of a hotel in Jamaica, allegedly flooded as a result of the storm. The second falsely portrayed the complete destruction of Kingston Airport. Analyses demonstrated that neither of these depicted events had actually occurred—both videos had been created entirely using the Sora system. The shark video was particularly notable for the speed of its dissemination, initially appearing with the watermark of a TikTok account identifying itself as a creator of AI-generated visual effects from the Dominican Republic, while the Instagram version circulated without any indication of its context or origin [5].

Within only a few hours, the videos had accumulated millions of views, appearing alongside authentic footage of actual damage, thereby causing significant confusion among users regarding the true extent of the disaster [10]. The situation became so pronounced that Jamaica's Minister of

Education, Dana Morris Dixon, publicly urged citizens to obtain information exclusively through official channels, stating that even she herself regularly received false content through numerous WhatsApp groups [5]. A similar pattern was observed in the United States, where emergency management officials in Louisiana warned that AI-generated content was making it increasingly difficult for the public to distinguish the actual consequences of the disaster from fabricated scenarios, such as alleged sharks swimming through the streets of flooded cities [11].

Experts in disinformation emphasize that the motivation behind the creation of such content rarely has an explicit political dimension; rather, it primarily reflects a conventional engagement-driven monetization model, in which content that provokes strong emotional reactions generates a greater number of clicks and views, regardless of its authenticity [5]. According to analysts, the case of Hurricane Melissa marked the point at which tools such as the Sora system became sufficiently accessible and convincing to enable the mass production of crisis-related disinformation without requiring technical expertise, thereby inaugurating a new and more dangerous phase in the relationship between synthetic media and crisis reporting [12].

CASE STUDY 2: THE LOS ANGELES WILDFIRES (JANUARY 2025)

The devastating wildfires that affected Los Angeles in January 2025—including the well-known Palisades Fire and the Sunset Fire in the Hollywood Hills—constitute the second key case study examined in this paper, as they illustrate how AI-generated content spreads in parallel with an actual disaster, further complicating the public's ability to distinguish authentic reporting from fabricated material. While the fires engulfed an area approximately twice the size of Manhattan Island, forcing nearly 180,000 people to evacuate, social media platforms were inundated with AI-generated images and videos within 24 hours of the onset of the disaster [6].

Among the most widely disseminated pieces of false content were videos depicting the iconic "Hollywood" sign allegedly completely surrounded by flames (Figure 2). In one such video, the word "Hollywood" was misspelled, providing a clear visual indication of its artificial nature [6]. Despite this, the video accumulated 1.3 million views on Facebook alone within 24 hours, while the official Hollywood Sign page was compelled to publicly refute the claims, confirming that the sign "is still standing" [13]. Analyses conducted by VERA Files, in cooperation with the Indian Deepfakes Analysis Unit, employed detection tools such as Hive AI, "Is it AI?", and "AI or Not" to confirm that all analyzed videos were "highly likely" or "extremely likely" to have been generated by artificial intelligence [7].



Figure 2. AI Generated Video Content During the Los Angeles Fires in 2025

In addition to the videos of the Hollywood Sign, other fabricated content also circulated—including footage allegedly showing firefighters rescuing animals from the flames—in which unnatural physical characteristics (distorted movements, inconsistent lighting, and background artifacts) clearly indicated AI generation. McAfee researchers traced the origin of many of these

videos to the Gemini tool, while some Instagram posts contained visible watermarks from tools such as Grok, which most users, despite their presence, failed to recognize as indicators of AI-generated content [6]. According to VERA Files, a single compilation of fabricated videos on Facebook alone accumulated more than 139,700 reactions, 18,900 comments, 58,400 shares, and 12.6 million views [7].

According to the fact-checking community, the Los Angeles wildfires are now regarded as a turning point at which AI-driven disinformation entered a qualitatively new and more dangerous phase—not only because of the scale of its dissemination, but also because the content was sufficiently convincing to deceive users who, due to their own exposure to danger and emotional distress, were less likely to rely on critical evaluation of the authenticity of the information they encountered [6].

AFP FACT CHECK: VERIFICATION OF AI-GENERATED CONTENT DURING NATURAL DISASTERS

What Is AFP Fact Check

AFP Fact Check (also known as **AFP Factuel** in its French-language editions) is the specialized fact-checking division of the French news agency *Agence France-Presse* (AFP), established in 2017 exclusively for the purpose of fact verification and combating disinformation [14]. Unlike individual independent fact-checking initiatives, AFP Fact Check operates within one of the world's oldest and largest news agencies, with teams distributed across the globe verifying content in 18 working languages, including French, English, Spanish, and Portuguese [14]. This linguistic and geographical diversity enables AFP to monitor and verify viral content in real time on a global scale, including in regions that are rarely covered by smaller, local fact-checking organizations.

Operating Principles

The methodology employed by AFP Fact Check is based on a combination of several standardized digital forensic techniques, which are consistently applied regardless of the platform or type of content being verified:

- Visual verification through reverse image search—using tools such as Google Reverse Image Search and the InVID-WeVerify extension, the team determines whether the disputed image or video has previously appeared online, in what context, and from which year it originates, thereby identifying cases in which older content has been recycled and presented under a false description [15].
- Geolocation — the disputed content is compared with satellite imagery and street-level images from Google Maps and Google Earth from the corresponding time period in order to determine whether the depicted location physically corresponds to the claimed place and time of recording.
- Consultation with independent sources and institutions — as a standard practice, AFP editors first attempt to contact the individual or institution that made the disputed claim, review AFP's own archives, and, where necessary, consult external experts in relevant specialized fields [16].
- Transparent publication — each fact-check report includes comparative visual evidence (the original and disputed images or videos displayed side by side), identifies the specific analytical tools used during the verification process, and, whenever possible, identifies the original creator of the content, thereby enabling readers to independently verify the presented conclusions [16].

This methodology was applied, among other cases, during Hurricane Melissa in October 2025, as well as in exposing the false claim regarding an alleged tsunami in Italy, where the combination of reverse image searching and geolocation enabled investigators to rapidly determine

that the depicted content did not correspond to the claimed time and location of the event. An example of an AFP report is shown in Figure 3.

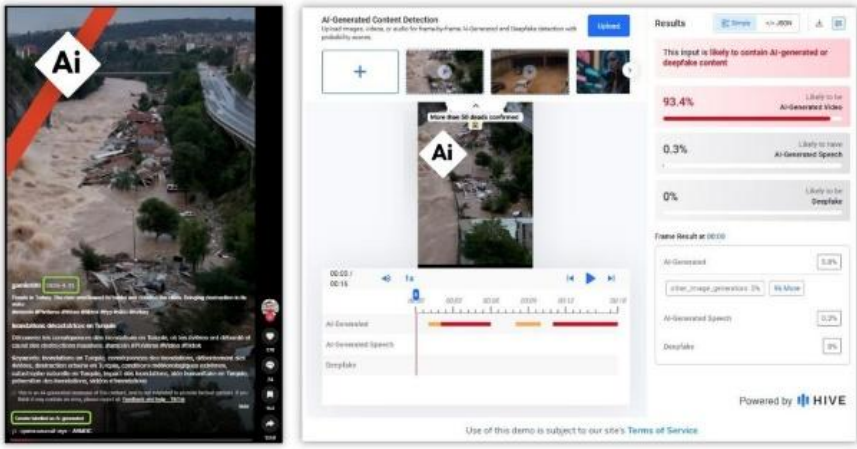


Figure 3. Screenshot of a TikTok post from 2025, annotated to display the publication date and the AI label (left), together with the results of a Hive Moderation analysis conducted on March 12, 2026. AI symbol added by AFP

How Credible Is AFP Fact Check?

The credibility of AFP Fact Check is derived from several independently verifiable factors:

- Institutional status and independence**.** AFP is legally governed by a special French law as an autonomous organization without shareholders. It is financed through the sale of its journalistic content to clients and by financial compensation from the French government solely for the costs associated with fulfilling legally defined public-interest missions. The law explicitly stipulates that AFP is neither a state-owned media organization nor affiliated with any political party [14].
- Independent external certification**.** AFP Fact Check is a certified signatory of the *International Fact-Checking Network* (IFCN) Code of Principles at the Poynter Institute. Consequently, its work is evaluated annually by independent external assessors against a set of 31 criteria, including impartiality, transparency of sources, methodology, and funding [17]. According to the most recent available assessment, more than 75% of AFP's fact-check reports address issues of public interest that may affect the well-being of individuals or society as a whole, thereby fulfilling the IFCN relevance criterion [14].
- Scale and continuity of operations**.** In 2020 alone, AFP Fact Check published more than 7,000 fact-check reports in 18 languages, demonstrating the scale and consistency required for reliable real-time monitoring of viral content [14].
- Platform partnerships**.** AFP is one of the independent fact-checking organizations engaged by Meta platforms (Facebook and Instagram) to verify content within their disinformation mitigation program. Consequently, AFP's labels identifying false content have a direct operational impact on the visibility of such content across these platforms [18].

Despite its high level of institutional credibility, it is important to acknowledge a limitation shared by all fact-checking services, including AFP: regardless of the methodological rigor employed, the verification process requires time—typically several hours to one or two days—whereas AI-generated content during crisis situations, as demonstrated in the preceding chapters, can achieve viral dissemination in less than 24 hours. This structural temporal asymmetry means that even the most credible fact-checking organizations generally operate reactively—exposing disinformation only after it has already reached millions of users—which raises the question

addressed in the following chapter concerning detection tools and the possibilities for early, proactive intervention.

DETECTION TOOLS

Why No Single Tool Is Sufficient on Its Own

Before presenting specific detection tools, it is important to emphasize a finding consistently confirmed by the academic literature: no individual algorithm or tool for detecting synthetic media can guarantee complete accuracy. A comprehensive systematic review published in *Discover Internet of Things* (Springer Nature) analyzed a wide range of detection algorithms—including convolutional neural networks (CNNs), long short-term memory (LSTM) models, transformer-based architectures, and metadata analysis—and concluded that the combination of audio analysis, facial analysis, and metadata examination constitutes a necessary prerequisite for reliable, rapid, and efficient deepfake detection. In contrast, reliance on a single detection technique has proven insufficient due to the rapid evolution of generative models and the limited availability of labeled training data [19]. This conclusion is directly supported by the practices described in the preceding chapters—organizations such as AFP, McAfee, and VERA Files systematically combine multiple tools (Hive Moderation, Attestiv, "Is it AI?", and "AI or Not") with reverse image searching and expert human assessment rather than relying on a single isolated result.

The Most Widely Used Commercial Detection Tools

The market for deepfake detection tools has experienced rapid growth in recent years, reaching an estimated value of approximately USD 581 million in 2025. Among the most frequently cited and widely adopted solutions are:

- Reality Defender—a multimodal detection platform (video, audio, image, and text) based on an ensemble approach that combines multiple models simultaneously. It is primarily used by financial institutions, government agencies, and enterprise security teams. At the end of 2025, the analytical firm Gartner identified Reality Defender as the leading company in the field of deepfake detection.
- Hive AI / Hive Moderation — a tool distinguished by reported detection accuracy exceeding 98% and the ability to identify outputs generated by leading content-generation systems, including Sora, Midjourney, Stable Diffusion, DALL·E, and Kling. Its capability to process large volumes of data makes it particularly suitable for content moderation on platforms serving millions of users. This was the primary tool employed in the analyses cited in the preceding chapters (AFP and McAfee). An example of this tool's report is shown in Figure 4.
- Sensity AI — with a reported detection accuracy of 95–98%, it is widely used by government agencies, cybersecurity companies, and media organizations for detecting face swaps, manipulated audio, and AI-generated images.
- Pindrop (Pindrop Pulse) — a solution specialized in audio deepfake detection, particularly in the context of voice-cloning fraud affecting call centers and financial services.
- Figure 5 presents video and image detection tools.

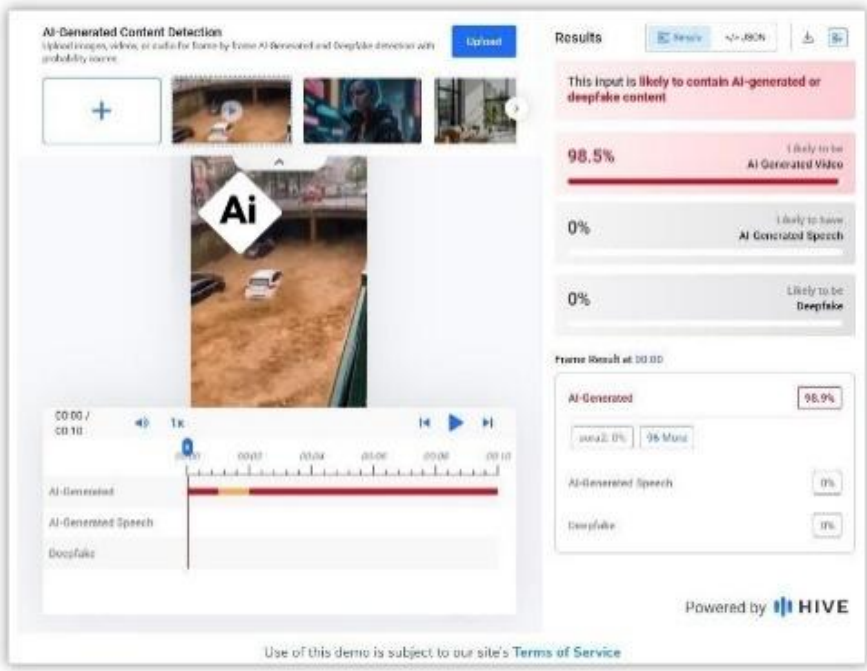


Figure 4. Screenshot of Hive Moderation analysis results, captured on March 12, 2026. AI symbol added by AFP

Video and Image Detection Tools

Tool	Type	Function / Description
InVID-WeVerify	Video / Frame	Frame-by-frame analysis, deepfake detection, and geolocation of video footage. Browser plugin developed with AFP Medialab.
Hive Moderation	Video / Image	Upload video or image to receive an AI-generation probability score. Fast automated screening.
Reality Defender	Video / Audio	Multimodal analysis of both video and audio content. Flags synthetic faces, voice, and scene inconsistencies.
Amped Authenticate	Video / Image	Forensic analysis of metadata, pixel structure, lighting anomalies, and DCT frequency spectrum.
Reverse Image Search	Image / Frame	Google Images / TinEye — traces the original source of video keyframes to expose reused or misattributed footage.

Figure 5. Video and image detection tool overview

It should be noted that the reported accuracy rates and market positions of these tools are derived primarily from technology industry reports and consulting analyses rather than independently peer-reviewed scientific studies. Consequently, these figures should be interpreted as indicative rather than absolute measures of reliability. Table 2 shows an overview of sound detection tools.

Table 2. Sound detection tool overview

Audio Detection Tools	Tool Type	Function / Description
ElevenLabs AI Speech Classifier	Audio	Detects whether an audio sample was generated by ElevenLabs' AI. Accessible via API; first tool of its kind publicly released (2023).
Resemble Detect	Audio	Real-time AI voice detection tool. Analyses audio samples for synthetic speech patterns and voice cloning artifacts.
AI Voice Detector (Hive)	Audio	Upload audio file to detect AI-generated speech. Outputs a probability score distinguishing human from synthetic voice.
Pindrop Pulse	Audio	Forensic audio analysis used in enterprise security contexts. Detects voice spoofing, deepfake audio, and liveness fraud.

Tools Used by Advanced Countries

At the governmental level, the most advanced and longest-running program in this field is led by the United States *Defense Advanced Research Projects Agency* (DARPA). The Media Forensics (MediFor) program, launched in 2016, and its successor, Semantic Forensics (SemaFor), launched in 2020, developed a comprehensive suite of forensic technologies capable of automatically detecting, attributing, and characterizing manipulated multimodal media content, including images, videos, audio recordings, and text [20]. Following the official completion of the SemaFor program in September 2024, DARPA continued transferring the developed technologies to U.S. government agencies and the commercial sector, while further research into media provenance and authenticity was assumed by the *Digital Safety Research Institute* (DSRI), part of *UL Research Institutes* [20]. This model—in which a government agency finances fundamental research and subsequently transfers the resulting technologies to academia, industry, and other governmental bodies through an open catalog of tools—appears, according to available information, to be followed by several other countries as well, although considerably fewer technical details are publicly available regarding their specific implementations.

Concluding Recommendation

Based on the findings presented, we consider it important to emphasize the following: responsible institutional practice during crisis situations should never rely solely on the output of a single detection tool, regardless of the level of accuracy claimed by its developers. Combining at least two or three independent methods—for example, algorithmic image or audio analysis (Hive, Sensity, or Pindrop), reverse image searching and geolocation (as employed by AFP and VERA Files), together with metadata and file provenance analysis (following the principles of the DARPA approach)—substantially reduces the risk of false-negative results, that is, situations in which synthetic content is incorrectly classified as authentic. Furthermore, since all commercial detection tools are based on models trained using previously known content-generation systems, their reliability naturally declines with each new generation of synthetic media generators (such as the Sora system, which, as demonstrated in the Hurricane Melissa case study, successfully bypassed many of the previously established indicators of manipulation). For this reason, even with the best technologies currently available, expert human judgment and institutional coordination—rather than automated detection alone—remain indispensable components of any reliable verification system during emergency situations.

CONCLUSION

This paper has demonstrated that AI-generated video and audio content represents one of the most significant contemporary challenges to crisis communication. The case studies of Hurricane Melissa and the Los Angeles wildfires confirm that false content can now reach millions of views in less than 24 hours—more rapidly than even the most credible fact-checking organizations, such as AFP, are able to complete the verification process. This dynamic is further exacerbated by algorithmic amplification, which systematically rewards emotionally engaging content regardless of its authenticity. The analysis of detection tools has shown that no single technical solution provides complete reliability and that the only dependable approach is the combination of multiple independent methods while maintaining expert human judgment.

A positive aspect of this development is that the same technological progress enabling the creation of increasingly convincing synthetic content is simultaneously driving the rapid advancement of technologies designed to detect it—from government-funded DARPA programs such as SemaFor, through commercial solutions such as Hive and Reality Defender, to regulatory frameworks including the EU AI Act, the Digital Services Act (DSA), and the C2PA initiative for verifying the provenance of digital content. Furthermore, the growing public awareness of the existence of deepfake content—as reflected in the user reactions described in the case studies—indicates a gradual strengthening of societal resilience to this form of manipulation.

The negative aspect, however, remains structurally more difficult to address. The authors of this paper believe that the quality of AI-generated content will continue to improve over the coming years to the point at which the distinction between authentic footage and synthetic media will no longer be recognizable, even to trained experts, let alone the general public—a trend already anticipated by the experts cited in this paper, who warn that telltale indicators such as distorted movements or misspelled signs are gradually disappearing with each new generation of generative models [5]. Should this trend continue at its current pace, detection technologies may permanently lag behind content-generation capabilities, causing the burden of verifying the truth to shift increasingly toward institutional and regulatory mechanisms rather than technology itself.

The consequences of this phenomenon extend far beyond mere informational confusion. During natural disasters, exposure to false content further burdens the psychological state of already vulnerable populations by intensifying fear, uncertainty, and distrust at precisely the moment when psychological stability is essential for responding appropriately to genuine danger. Moreover, the increasingly prominent phenomenon known as the *liar's dividend*—where awareness of the existence of deepfake content leads individuals to dismiss authentic videos and legitimate warnings as fabricated—further erodes public trust in official sources, even when those sources are conveying accurate information. This effect is particularly dangerous during crisis situations, where hesitation in responding to genuine warnings may result in the loss of life.

In the authors' view, the most serious operational risk concerns the work of emergency response and rescue services. When fabricated or incorrectly geolocated content—for example, footage of a "collapsed bridge" or a "flooded neighborhood" that actually originates from a different location or time—is indiscriminately shared and perceived as credible, emergency responders may be redirected to the wrong location while the actual disaster unfolds elsewhere, thereby losing precious time during which lives could otherwise be saved. This scenario is not merely hypothetical: following the earthquake in northwestern England in 2025, Network Rail suspended railway traffic in the affected area on the basis of a fabricated photograph allegedly depicting a collapsed bridge, illustrating how rapidly synthetic content can produce real operational consequences in the field [21].

Based on the findings presented, it can be concluded that effectively countering AI-generated disinformation during crisis situations requires a comprehensive, multi-layered approach combining: (1) the continuous improvement and integration of detection technologies; (2) clear regulatory requirements mandating the labeling of AI-generated content; (3) proactive and timely institutional communication capable of preventing information vacuums; and (4) long-term investment in public digital literacy. Without such an integrated response, the boundary between real and artificially generated disasters risks becoming increasingly difficult to distinguish, with direct consequences for

public safety, the psychological resilience of affected populations, and the effectiveness of emergency response efforts.

LITERATURA

1. Vosoughi, Roy & Aral, 2018: https://doi.org/10.1126/science.aap9559
2. RAND Corporation: https://www.rand.org/research/projects/truth-decay.html
3. JMIR Infodemiology, 2025: https://infodemiology.jmir.org/2025/1/e70413
4. Logically AI, 2026: https://logically.ai/case-studies/case-study-weaponized-weather-when-disasters-become-information-battlegrounds
5. PBS NewsHour: https://www.pbs.org/newshour/world/phony-ai-videos-of-hurricane-melissa-flood-social-media
6. McAfee, 2025: https://www.mcafee.com/blogs/internet-security/how-social-media-is-spreading-l-a-misinformation-like-wildfire/;
7. VERA Files, 2025: https://verafiles.org/articles/fact-check-ai-generated-videos-go-viral-as-california-fires-rage-on
8. PBS NewsHour, 2025: https://www.pbs.org/newshour/world/phony-ai-videos-of-hurricane-melissa-flood-social-media
9. Logically AI, 2026: https://logically.ai/case-studies/case-study-weaponized-weather-when-disasters-become-information-battlegrounds
10. France24, 2025: https://www.france24.com/en/tv-shows/truth-or-fake/20251029-ai-generated-videos-of-hurricane-melissa-flood-social-media
11. NOLA.com, 2025: https://www.nola.com/news/weather/hurricane-melissa-ai-content-louisiana/article_4f99fc35-ed65-4218-8650-8ddc5546ea2f.amp.html
12. Boston Globe, 2025: https://www.bostonglobe.com/2025/10/30/business/phony-ai-videos-hurricane-melissa/
13. Deadline, 2025: https://deadline.com/2025/01/the-hollywood-sign-not-on-fire-social-media-reacts-ai-generated-images-1236252934/
14. IFCN Code of Principles, Poynter: https://ifcncodeofprinciples.poynter.org/application/public/afp-fact-checking/CF0D33C4-39AC-B1EE-7057-4B35A5CDE953
15. Chulalongkorn University/AFP Digital Courses: https://www.chula.ac.th/en/highlight/272509/
16. IFCN Code of Principles, Poynter: https://ifcncodeofprinciples.poynter.org/application/public/afp-fact-checking/160AE4ED-E09A-FA64-42F0-3D5298F00A62
17. IFCN, Poynter: https://ifcncodeofprinciples.poynter.org/profile/afp-fact-checking
18. Wikipedia — Lista fact-checking sajtova: https://en.wikipedia.org/wiki/List_of_fact-checking_websites
19. Springer Nature, *Discover Internet of Things*, 2025: https://link.springer.com/article/10.1007/s43926-025-00154-0
20. DARPA, 2025: https://www.darpa.mil/news/2025/furthering-deepfake-defenses
21. TechPolicy.Press, 2026: https://www.techpolicy.press/ai-detection-was-built-for-faces-climate-deception-targets-environments/
22. trini.man47/TikTok; zakirtiussain5988/TikTok