

The Crisis of Digital Authenticity in the Age of AI-Generated Content and New Media

Bo Jin

*Media and Communication, Monash University, Melbourne, Australia
bjin0013@student.monash.edu*

Abstract. The widespread adoption of generative artificial intelligence (AI) technologies has reduced the technical barriers to producing fabricated images, videos, and voice recordings. The prioritization of high-engagement content by recommendation algorithms on major new media platforms allows AI-generated content to receive greater exposure within digital information systems. Simultaneously, the mere existence of AI has weakened public trust in authentic content, as genuine news photographs and videos can increasingly be dismissed as "AI-generated." This phenomenon, known as the "liar's dividend," weakens trust in digital evidence. This study analyzes the generative mechanisms and social consequences of this authenticity crisis through three dimensions: technological barriers, platform amplification, and public trust. Using literature analysis and case-based evidence, it examines the broader implications of AI-generated content for journalism and public trust systems. Therefore, the results reveal that AI-generated content has transformed perceptions of digital authenticity, creating challenges for public trust and news credibility in the new media era.

Keywords: Artificial Intelligence, AI-Generated Content, Digital Authenticity, Synthetic Media, Trust Crisis

1. Introduction

As artificial intelligence (AI) technologies become more widely used in digital media creation and distribution, tools such as image generation, voice cloning, and AI-generated video have enabled the production of increasingly realistic synthetic content [1]. Existing research has largely focused on the technical capabilities of deepfake generation and the risks of synthetic media [2]. However, less attention has been given to how AI-generated content transforms the standards through which audiences evaluate authenticity, credibility, and digital evidence. Rather than simply increasing the amount of false information, AI technologies may alter the broader information environment by changing how individuals, media organizations, and institutions determine what can be considered trustworthy. This paper examines how AI-generated content shapes the production, dissemination, and evaluation of digital information. Moreover, it focuses particularly on the changing standards of authenticity judgment, the increasing ambiguity between authentic and fabricated content, and the decline of trust in the digital information environment. Through literature analysis and case studies, it explores the challenges created by AI-generated media and considers their implications for media governance, journalistic credibility, and public media literacy.

2. The shift in false content production barriers driven by artificial intelligence

2.1. The emergence of AI-generated content creation tools

Generative AI has shifted synthetic media production from editing existing materials to generating new content through automated models. This development has led to the emergence of tools such as Google's Gemini and OpenAI's Sora, which can generate highly realistic images and videos from text prompts. Meanwhile, advanced voice cloning technologies can replicate human speech patterns [3, 4]. Earlier digital manipulation techniques mainly altered existing media, whereas generative AI systems can create entirely new multi-modal content with greater realism and complexity. This shift represents a transition from manual media modification to AI-assisted content creation, expanding the capabilities of digital media production.

2.2. The advancement of synthetic media and deepfake technologies

Although generative AI tools can generate various types of synthetic media, deepfake technologies specifically aim to imitate real individuals by reproducing their identities, appearances, voices, and behaviors. More specifically, deepfakes are designed to imitate specific people with high levels of realism, thus increasing the potential for deception and misuse. As these technologies become more accessible and sophisticated, their applications have expanded beyond entertainment and personal manipulation into high-risk domains such as legal and financial systems. For example, AI-generated video evidence has raised concerns about the reliability of digital recordings in legal proceedings, as illustrated by *Mendones v. Cushman & Wakefield*, where an AI-generated video was submitted as part of the legal proceedings [5, 6]. The increasing ability to fabricate convincing audio and visual materials challenges existing methods of verification and creates broader concerns about trust in digital evidence. Furthermore, the continued development of deepfake technologies has intensified concerns about their potential to undermine legal, financial, and institutional systems [1].

2.3. The reduction of costs and barriers in false content production

The increasing accessibility of AI tools has significantly lowered the technical and financial barriers associated with producing inauthentic content. Specifically, tasks that previously required advanced editing skills, specialized software, and extensive resources can now be completed with readily available consumer technologies [2]. As a result, a wider range of users can create realistic synthetic images, audio, and videos with limited technical expertise. This reduction in production barriers enables the faster creation and wider distribution of fabricated content, increasing the scale and frequency of synthetic media production. The ability to discredit authentic content by claiming it is fabricated has been described as the "liar's dividend," hence highlighting a potential consequence of synthetic media that will be explored in later sections [1].

3. The dissemination of AI-generated content in the new media platform environment

3.1. Algorithmic recommendation and content spread

On new media platforms, algorithmic systems organize and prioritize content by determining the level of exposure each item receives [7]. These systems analyze various engagement-related signals, including clicks, shares, comments, viewing duration, and interaction rates, to rank and recommend

content to users. Rather than independently assessing the accuracy or authenticity of information, recommendation systems primarily optimize for user engagement and retention.

The operation of these algorithms involves selecting, prioritizing, and organizing large volumes of content based on predicted user interests and platform objectives. As a result, content visibility is increasingly determined by algorithmic ranking processes rather than solely by chronological order or editorial decisions. As generative AI advances, the expansion of synthetic content increases the role of algorithmic systems in filtering, ranking, and presenting digital materials to users [7].

3.2. Platform engagement and audience response

The interaction between AI-generated content and audience responses plays an important role in shaping the dissemination of synthetic media. In this context, users' engagement with such content is often influenced by the perceived realism and credibility of the generated materials. When this realism reaches a high level, users may experience difficulties in distinguishing fabricated content from authentic materials [1, 8].

Existing studies have shown that individuals often struggle to distinguish AI-generated imagery from authentic content, particularly when synthetic materials closely resemble real media [1, 8]. In these situations, users may rely on perceived credibility rather than the actual origin of the content when deciding whether to engage with or share it. Consequently, AI-generated materials that appear authentic may receive user interactions and circulate through social media platforms before their synthetic nature is identified. Through this process, AI-generated content can spread through user interactions before its origin and authenticity are confirmed. With the continued improvement and accessibility of synthetic media, audience responses have become an important factor influencing how such content is distributed across digital platforms.

3.3. Information consumption and user verification behaviors

Fragmented information consumption has accelerated the circulation of AI-generated content across digital platforms. On digital platforms, users often engage with headlines, images, and short video clips rather than reading complete articles or detailed reports. And this rapid and selective form of consumption encourages immediate sharing and redistribution, allowing digital materials to move across social networks and private communication channels within a short period of time. During major news events, previously circulated materials and newly emerging content can be repeatedly edited, combined, and redistributed, enabling similar materials to reach different audiences across multiple platforms [9].

User engagement and platform distribution mechanisms further accelerate the spread of synthetic media. Interactions such as sharing, commenting, and reposting can increase the visibility of content and facilitate its movement across interconnected networks. Existing research has found that false news on Twitter achieves wider and faster dissemination than accurate information [7]. Although their study was conducted before the widespread adoption of generative AI, the findings reveal how engagement-driven communication systems can accelerate the spread of attention-grabbing content, providing insight into the potential amplification of AI-generated materials.

4. The impact of AI-generated content on perceptions of authenticity and social trust

4.1. The evolution of authenticity judgment standards

The emergence of AI-generated content has altered the way audiences determine whether digital information is authentic. Traditionally, audiences often relied on visual and audio-visual evidence as main indicators of credibility, assuming that realistic images, videos, or recordings were closely connected to real events. This reliance was reflected in the common assumption that "seeing is believing" [1, 9]. However, realistic AI-generated images, videos, and audio have weakened the link between realism and authenticity. Thus, visual plausibility alone is no longer sufficient to establish whether digital content is genuine.

This shift has changed how authenticity is cognitively assessed. Audiences must now consider additional factors beyond the appearance or quality of media, including the context in which content is produced, distributed, and interpreted. The presence of highly realistic synthetic media has also increased uncertainty in authenticity judgments, as realistic content may no longer provide reliable evidence of its origin. Moreover, confirmation bias may influence evaluations of politically related content, causing individuals to interpret information through existing beliefs rather than objective criteria [9, 10]. AI-generated media therefore does not merely create individual cases of deception; it changes the broader standards through which digital evidence is understood and evaluated.

4.2. The blurring of the boundary between authentic and fabricated content

By generating realistic forgeries and enabling challenges to genuine evidence, AI-generated content has blurred the boundary between authentic and fabricated media. While synthetic media can make fabricated materials appear increasingly realistic, the widespread availability of AI generation tools also provides a basis for questioning the authenticity of genuine evidence. Thus, visual realism is no longer a sufficient indicator of credibility, and digital materials can no longer be evaluated solely through their appearance. The "liar's dividend" represents one consequence of this transformation, in which authentic content can be rejected by attributing it to AI manipulation. Beyond producing false materials, AI-generated media reshapes the criteria used to evaluate the credibility of digital evidence.

For instance, in 2024, Donald Trump questioned the authenticity of a verified photograph with E. Jean Carroll by claiming that it was AI-generated. Rather than being used to produce false content, AI technology provided a mechanism for challenging genuine evidence, weakening the evidentiary value of digital records [10]. Additionally, in 2025, Trump questioned the authenticity of security footage showing an object falling from an upper window of the White House, despite confirmation that the footage was genuine. These cases reveal that AI-related claims can weaken confidence in authentic evidence by making digital materials appear uncertain. Therefore, AI-generated content extends beyond the creation of fabricated media by blurring the boundary between authentic and unreliable evidence, complicating digital verification [1, 2].

4.3. The decline of trust in the digital information environment

The increasing presence of AI-generated content has contributed to declining trust in news media and the broader digital information environment. As authentic and synthetic materials increasingly appear within the same information spaces, the traditional relationship between published evidence and perceived credibility becomes less stable. Thus, audiences may experience greater uncertainty

regarding the reliability of digital materials, while news organizations face greater challenges in verifying reported information. This challenge is illustrated by early January 2026 events, where AI-manipulated images circulated alongside legitimate materials, weakening confidence in digital information reliability [9]. At a broader level, news organizations face increasing challenges in establishing a clear hierarchy of trustworthy sources, as verified reporting competes with rapidly circulating synthetic materials within the same media environment [11].

This decline in trust extends beyond individual cases of misinformation and influences broader engagement with public information [12, 13]. When audiences increasingly perceive digital content as potentially unreliable, the value of news media as a source of shared knowledge may gradually diminish. This concern is reflected in survey findings, as a survey conducted in August 2024 found that around 50% of American adults expressed concern about the growing presence of AI in daily life. Reduced confidence in digital information can affect how individuals engage with public issues, potentially limiting informed discussion and collective decision-making. The broader consequence of AI-generated misinformation lies in the erosion of shared trust in the information environment, which supports institutional accountability, public discourse, and social coordination [1, 2, 14].

5. Conclusion

This study indicates that AI-generated content has transformed the conditions under which digital information is produced, distributed, and evaluated. The impact of AI extends beyond the creation of fabricated media. In particular, generative AI technologies have lowered the technical barriers to producing realistic synthetic content, increasing the accessibility of media manipulation. Meanwhile, platform-based recommendation systems can amplify AI-generated content by prioritizing highly engaging materials. Moreover, AI-generated content has changed how audiences assess authenticity. While synthetic media can make fabricated materials appear credible, AI also enables authentic evidence to be questioned as manipulated, contributing to a broader trust crisis.

For journalism, this shift reshapes the challenges of maintaining information credibility beyond traditional misinformation control. Thus, news organizations must strengthen verification practices and develop clearer methods for establishing the authenticity of digital evidence. Public trust relies on verification systems that identify fabricated content while preventing legitimate information from being dismissed as manipulated. As the boundary between authentic and synthetic media becomes less distinct, journalism must improve transparency mechanisms and media literacy efforts to maintain credibility. This study is primarily based on literature analysis and selected case studies, which limits the diversity of contexts examined. Future research should examine diverse countries, platforms, and media environments through comparative analysis and audience studies. It should also evaluate whether content provenance systems and AI-traceability tools can improve trust in digital information.

References

- [1] Chesney, B., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *Calif. L. Rev.*, 107, 1753.
- [2] Goldstein, J. A., & Lohn, A. (2024). Deepfakes, elections, and shrinking the liar's dividend. Brennan Center.
- [3] Wan, T., Wang, A., Ai, B., *et al.* (2025). Wan: Open and advanced large-scale video generative models. arXiv preprint arXiv:2503.20314.
- [4] Jin, C., Fan, J., Huang, J., *et al.* (2026). Next-gen AIGC: A review of multimodal foundation models for text-to-media innovations. *Frontiers of Computer Science*, 20(12), 2012368.

- [5] Matthewman, W. (2026). The coming use and misuse of artificial intelligence in the courtroom: A judicial perspective and proposal. *Journal of Technology Law & Policy*, 30(2), 1.
- [6] Lao, Y., Hirvonen, N., & Larsson, S. (2026). AI and authenticity: Young people's practices of information credibility assessment of AI-generated video content. *Journal of Information Science*, 52(3), 975-989.
- [7] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.
- [8] Hamilton, H. (2026). Machines spot deepfake pictures better than humans, but people outperform AI in detecting deepfake videos. <https://news.ufl.edu/2026/02/deepfake-detection/>
- [9] Yang, A. (2026). AI is intensifying a 'collapse' of trust online, experts say. <https://www.nbcnews.com/tech/tech-news/experts-warn-collapse-trust-online-ai-deepfakes-venezuela-rcna252472>
- [10] Lee, J., Booth, E., Farid, H., & Rizoio, M. A. (2025). Misinformation is not about bad facts: An analysis of the production and consumption of fringe content. *EPJ Data Science*, 14(1), 50.
- [11] Yanes, N. A., Jitkajornwanich, K., Kee, K. F., & Perlmutter, D. D. (2026). The original sins of algorithmic authenticity verification: Provenance neglect and model collapse in AI-mediated news and information. *Journal of Broadcasting & Electronic Media*, 1-21.
- [12] Madrid-Morales, D., Wasserman, H., Gondwe, G., et al. (2021). Motivations for sharing misinformation: A comparative study in six Sub-Saharan African countries. *International Journal of Communication*, 15(2021), 1200-1219.
- [13] Romanishyn, A., Malytska, O., & Goncharuk, V. (2025). AI-driven disinformation: Policy recommendations for democratic resilience. *Frontiers in Artificial Intelligence*, 8, 1569115.
- [14] Schiff, D. S., Schiff, K. J., & Bueno, N. (2024). Watch out for false claims of deepfakes, and actual deepfakes, this election year.