RESEARCH ARTICLE

2023, vol. 10, issue 2, 330 - 335
https://doi.org/10.5281/zenodo.

The ethics of AI-generated imagery in journalism

Dan Valeriu VOINEA

University of Craiova, Romania

Abstract

The rapid advancement of generative artificial intelligence (AI) capable of producing highly realistic images presents profound ethical challenges for journalism. This paper examines the multifaceted ethical landscape surrounding AI-generated imagery in news contexts, synthesizing current scholarship and expert perspectives. It analyzes the significant risks posed by synthetic media, focusing on four key themes: the potential for AI-generated visuals to propagate misinformation and erode public trust; the direct challenges to journalistic integrity and traditional ethical codes demanding truth and accuracy; the complex dynamics of audience perception, media literacy, and trust in an era where distinguishing authentic visuals from fakes is increasingly difficult; and the development of policy recommendations and best practices for navigating this terrain. The paper highlights the urgent need for rigorous verification, clear labeling and disclosure, restricted use in factual reporting, updated internal policies, leveraging authentication technologies, and a continued commitment to core journalistic values to maintain credibility and public trust amidst the transformative impact of AI on visual information.

Keywords: AI-Generated Imagery; Journalism Ethics; Misinformation; Public Trust; Synthetic Media; Transparency;

Introduction

The rapid advancement of generative artificial intelligence (AI) presents profound ethical challenges for journalism, a profession fundamentally grounded in principles of truth, accuracy, and public trust. AI systems can now create highly realistic images and videos – often hard to tell apart from real photos or footage – which are spreading rapidly online and through media channels (Arguedas & Simon, 2023; Bendel, 2023). High-profile instances, such as fabricated images of political events or public figures, underscore the potential for AI-generated visuals to mislead audiences on a significant scale (Jones et al., 2023). As news organizations navigate the complexities of using, identifying, or debunking synthetic media, critical questions emerge regarding the impact on misinformation, the preservation of journalistic integrity, audience interpretation, and trust. Alongside these risks, discussions also involve exploring potential responsible applications, such as data visualization or illustration, demanding necessary ethical guidelines for practice in this new technological era..

This paper examines the multifaceted ethical landscape surrounding AI-generated imagery within journalism. It synthesizes current scholarship and expert perspectives to analyze four central themes: (1) the role of AI-generated visuals in propagating misinformation and eroding public trust; (2) the challenges posed to journalistic integrity and traditional ethical codes; (3) the dynamics of audience perception, media literacy, and trust in the age of synthetic media; and (4) the development of policy recommendations and best practices for navigating this complex terrain. By exploring existing research, identifying patterns, and highlighting gaps in current understanding, this paper aims to illuminate the significant risks and responsibilities associated with AI-generated imagery and chart a course for maintaining journalistic credibility.

AI-Generated Imagery and Misinformation

One of the most significant ethical concerns arising from AI-generated visuals is their potent capacity to create and disseminate misinformation. Synthetic images and deepfake videos can convincingly depict events that never occurred or subtly alter real events, thereby deceiving audiences and potentially manipulating public opinion (Jones et al., 2023; Vyas, 2022). The creation of hyper-realistic but entirely fabricated scenes, such as fictitious political scandals or non-existent disasters, fundamentally challenges the adage that "seeing is believing" (Habgood-Coote, 2023). Imagine, for instance, a convincingly generated image depicting a politician secretly

meeting with a foreign adversary, or a fabricated video showing emergency services responding to a disaster that never occurred – such visuals could easily mislead the public if presented as fact.

Image 1. "Donald Trump, Vladimir Putin and Xi Jinping discussing splitting europe." Prompt, generated with ChatGPT.

Research indicates that deepfakes blur the lines between reality and fabrication, threatening to undermine a shared understanding of events and fostering uncertainty about the veracity of all visual media. This erosion applies not only to manipulated images but also casts doubt on genuine news imagery, as the mere possibility of undetectable fakes diminishes overall credibility (Bendel, 2023).

The proliferation of AI-manipulated visuals directly fuels the spread of disinformation. Malicious actors can leverage these tools to create compelling fake evidence supporting false narratives, which can rapidly gain traction when amplified through social media platforms (Jones et al., 2023; Vyas, 2022). The speed at which such content can go viral often outpaces the efforts of fact-checkers, allowing misinformation to take root before corrections can be widely disseminated (Arguedas & Simon, 2023). Public trust is particularly vulnerable when AI-generated images appear photorealistic and lack context, leading audiences to accept them as authentic, especially if they align with pre-existing biases.

Beyond direct deception, the existence of sophisticated AI fakes introduces the pernicious effect known as the "liar's dividend" (Chesney & Citron, 2019). This phenomenon describes the advantage gained by dishonest actors when public awareness of convincing fakes leads to skepticism toward even authentic evidence. Public figures facing genuine accusations can exploit this doubt by dismissing real photographic or video evidence as "deepfakes," thereby evading accountability (Schiff & Schiff, 2023). This strategic manipulation of public uncertainty undermines evidence-based journalism and poses a significant threat to democratic processes that rely on factual information (Arguedas & Simon, 2023). The potential for AI-fueled misinformation is now recognized as a major global risk, highlighting the scale and severity of this challenge (Jones et al., 2023).

In response, journalists face the dual burden of debunking AI-generated fakes while simultaneously avoiding the unwitting amplification of such content. News organizations are increasingly incorporating forensic techniques and AI-detection tools to authenticate visuals, though these methods face limitations in an ongoing technological arms race (Jones et al., 2023). The ethical mandate is clear: rigorous verification is paramount, and unverified visuals should not be published. Furthermore, journalists have a responsibility to inform the public about

significant circulating fakes. However, fulfilling this duty demands resources and expertise that may be scarce in fast-paced news environments, placing considerable strain on journalistic processes (Demmar & Neff, 2023).

Challenges to Journalistic Integrity

The emergence of AI-generated imagery directly confronts the core tenets of journalistic integrity, particularly the commitment to truthfulness and accuracy in reporting. Visual journalism and photojournalism have historically relied on the perceived factuality of images as records of reality (Gortázar, 2023 - Note: Source excluded as per date constraint). Established ethical codes, such as those from the National Press Photographers Association (NPPA) and the Society of Professional Journalists (SPJ), explicitly prohibit manipulations that distort truth (Paik et al., 2023). AI-generated or significantly altered images inherently conflict with these principles, as they represent manufactured content rather than authentic documentation of events (Bendel, 2023; Paik et al., 2023).

Journalists are thus presented with a significant ethical dilemma regarding the use of AI tools for image creation or alteration without compromising audience trust (Liu et al., 2023). While potential applications exist for illustrative purposes (e.g., visualizing abstract concepts or accompanying satire), there is a risk of blurring the lines between authentic and synthetic visuals in the audience's perception (Jones et al., 2023). The intent behind using an AI image is critical; deceptive intent is a clear violation of integrity. However, even unintentional manipulation or a lack of transparency regarding an image's synthetic origin can mislead viewers, as demonstrated by incidents where AI tools inadvertently altered the meaning of news photos.

Traditional journalistic codes offer relevant guidance, emphasizing truth, accuracy, and disclosure of manipulations likely to mislead (MEAA, 2019; SPJ Code of Ethics). The challenge lies in applying these longstanding principles to the novel capabilities of AI. Publishing an AI-generated image as if it were a real news photograph constitutes a severe ethical breach, akin to fabrication (Jones et al., 2023). Even omitting disclosure, allowing audiences to assume authenticity, represents deception by omission. Ethical practice demands transparency: AI-generated content must never be presented, implicitly or explicitly, as factual documentation of reality.

Within the journalism community, there is considerable unease about integrating synthetic content into news reports. Some professionals advocate for a strict prohibition, arguing that AI imagery is fundamentally incompatible with the principles of photojournalism, which derives its value from authentic representation (Palmer, 2023). This perspective highlights a perceived existential threat to journalism's core mission if the line between captured reality and generated fiction becomes blurred (Amanbay, 2023). While understandable, this purist stance risks overlooking potential ethical uses for illustration or data visualization, suggesting the need for nuanced guidelines rather than outright bans.

In response, news organizations are beginning to formulate policies. Many major outlets have prohibited or severely restricted the use of AI-generated visuals in editorial content, particularly photorealistic images intended to illustrate news events (Bauder, 2023; Jones et al., 2023). Guidelines often stipulate that AI-generated images should only be used when the AI generation itself is the subject of the story (AP, 2023). This approach aims to preserve integrity by ensuring audiences can trust that news visuals represent reality.

However, gray areas persist. Using AI for reconstructions or hypothetical illustrations requires extreme transparency to avoid misrepresentation (Jones et al., 2023). Furthermore, the inherent biases within AI models present another ethical challenge. Generative systems trained on biased data can produce stereotypical or distorted representations, potentially undermining fairness and accuracy in visual reporting (Bendel, 2023; Sarhan & Hegelich, 2023). Using such images could inadvertently introduce bias, violating the journalistic duty to minimize harm and report fairly.

Finally, there is a recognized gap and a need for updated ethical guidelines and training within the profession (Demmar & Neff, 2023). Some journalists express concern that ethical frameworks are not evolving quickly enough to address the rapid advancements in AI technology. Maintaining integrity requires both a recommitment to core values and proactive efforts to develop clear policies, provide adequate training, and maintain open communication with the audience about image production methods (Arguedas & Simon, 2023).

Audience Perception and Media Literacy

Understanding how audiences perceive and react to AI-generated visuals is critical to assessing their ethical impact. Visuals are powerful communication tools, capable of evoking strong emotions and influencing beliefs (Sarhan & Hegelich, 2023). A primary concern with AI imagery is the potential for audiences to be deceived by realistic fakes or, conversely, to develop excessive skepticism towards all visual media upon learning that convincing fakes are possible (Jones et al., 2023). Both outcomes significantly affect trust in journalism.

Initial research and anecdotal evidence suggest that many individuals struggle to distinguish sophisticated AI-generated images from authentic photographs (Arguedas & Simon, 2023). The inherent tendency to trust visual

information means that well-crafted fakes presented in news-like contexts can easily mislead viewers (Habgood-Coote, 2023). Visual misinformation can thus gain traction quickly, often before corrections or clarifications reach the public, exploiting a lag in audience awareness and verification habits (Jones et al., 2023).

As public awareness of deepfakes and AI manipulation grows, however, a counter-reaction of distrust can emerge. Knowledge that visual media can be convincingly fabricated may lead viewers to question the authenticity of legitimate news visuals, fostering a generalized skepticism (Habgood-Coote, 2023). While historical precedents with technologies like photo manipulation suggest society can adapt without entirely rejecting visual evidence (Habgood-Coote, 2023), the current period is marked by heightened uncertainty regarding the trustworthiness of news imagery (Arguedas & Simon, 2023).

Media literacy research offers insights into how labeling and disclosure affect audience interpretation. Studies indicate that clear labels identifying content as "manipulated" or "deepfake" effectively signal deception to audiences, reducing belief and sharing intentions (Eastwood, 2023). Neutral labels like "AI-generated," however, may not inherently convey trustworthiness or intent without further context (Eastwood, 2023). This suggests that transparent labeling by news organizations can mitigate some negative effects by prompting audience skepticism (Jones et al., 2023).

Effective labeling, however, presents challenges. Terminology must be widely understood, and inconsistent application can reduce effectiveness (Eastwood, 2023). Furthermore, the prevalence of labels might inadvertently lead audiences to assume unlabeled content is authentic, which is problematic if labeling is not comprehensive (Eastwood, 2023). Pairing labels with educational context could enhance audience understanding (Demmar & Neff, 2023).

The psychological impact also includes emotional and cognitive responses. Preliminary studies comparing human-selected news photos with AI-generated counterparts suggest potential differences in perceived quality and emotional tone, with AI images sometimes judged as less informative or contextually nuanced, though still capable of evoking strong emotions (Paik et al., 2023). Learning that an emotionally resonant image was fabricated could foster cynicism and betray audience trust (Bendel, 2023).

Ultimately, trust in journalism is at stake. Transparency consistently correlates with trust; audiences value openness about sourcing and methods (Arguedas & Simon, 2023). Suspicions of undisclosed AI use can erode fragile trust, while clear stances and consistent practices regarding AI imagery can reinforce it (Jones et al., 2023). Public concern about deepfakes creates an opportunity for news organizations to demonstrate leadership in combating visual misinformation. Enhancing media literacy through educational initiatives, teaching critical evaluation skills (like reverse image searching and source verification), can empower audiences to navigate the complex visual landscape (Demmar & Neff, 2023). Trust varies across demographics and cultures, necessitating context-sensitive transparency practices (Eastwood, 2023). Consistent honesty and clear expectation-setting are crucial for maintaining long-term audience trust in an era of synthetic media (Jones et al., 2023).

Policy Recommendations and Best Practices

Addressing the ethical complexities of AI-generated imagery necessitates the development and implementation of clear policies and best practices at multiple levels: within individual news organizations, across the journalism industry, and within the broader technological and regulatory environment (Jones et al., 2023).

1. Rigorous Verification and Source Transparency: News organizations must maintain the highest standards of verification for all visual content, treating AI-generated material with extreme skepticism (Bauder, 2023). Images lacking verifiable authenticity should not be used as factual representations of news events. Implementing forensic analysis protocols, even if they slow publication, is preferable to risking the dissemination of fakes (Jones et al., 2023). Documenting and disclosing image provenance whenever possible enhances transparency and audience trust (Sarhan & Hegelich, 2023).

2. Clear Labeling and Disclosure: When AI-generated imagery is used (e.g., for illustrative purposes), it must be explicitly and unambiguously labeled using clear, audience-tested language (Eastwood, 2023; Jones et al., 2023). Labels should be prominent and prevent any misinterpretation of the image as a factual depiction. Disclosure can be reinforced within accompanying text. Internal labeling protocols are also crucial to prevent accidental publication without context (Bendel, 2023).

3. Restricted Use in Factual Reporting: A consensus is emerging that AI-generated visuals should not substitute for photojournalism in depicting actual news events (AP, 2023; Jones et al., 2023). Real photographs and videos should always take precedence. Use in non-factual contexts (e.g., abstract illustrations, opinion pieces) requires careful consideration and extreme transparency. A clear distinction between reality and AI-generated fiction in news visuals must be maintained (Arguedas & Simon, 2023).

4. Internal Policies and Training: News organizations urgently need comprehensive, written AI ethics policies covering permissible uses, review processes, handling third-party content, and corrections (Jones et al., 2023). These policies should be developed consultatively, involving editorial staff, visual journalists, ethicists, and legal counsel (Liu et al., 2023). Regular updates and robust training programs are essential to ensure newsroom staff understand AI technologies, their pitfalls, ethical implications, and detection methods (Demmar & Neff, 2023). Publicly sharing these policies can further enhance transparency and trust (Jones et al., 2023).

5. Leveraging Authentication Technologies: The industry should actively engage with and adopt emerging technical solutions for content authentication, such as Content Credentials (CAI/C2PA) which embed metadata about an image's origin and manipulation history (Jones et al., 2023; Arguedas & Simon, 2023). While not foolproof, integrating these tools, alongside AI detection algorithms and potential watermarking standards, provides layers of defense against fakes. Collaboration between media, technology firms, and researchers is vital for developing reliable authentication systems (Brookings Institution, 2023).

6. Ethical Exploration of Beneficial Uses: While mitigating risks is paramount, policies should also consider potential beneficial applications of AI that do not compromise truth, such as assisting with data visualization, creating non-photorealistic illustrations, or restoring archival images (Jones et al., 2023; Liu et al., 2023). Such uses require clear rules and labeling to ensure they serve an explanatory function and cannot be mistaken for factual reporting. Strategic, safeguarded use may be more sustainable than blanket bans (Bendel, 2023; Jones et al., 2023).

7. Broader Regulatory and Industry Action: Addressing malicious deepfakes and large-scale disinformation may require legislative action, such as laws criminalizing harmful fabrications and mandating disclosure for AI-generated political content (Jones et al., 2023; Arguedas & Simon, 2023). Journalists and media organizations should engage in these policy debates. Fostering industry consortia to share information, develop standards, and advocate for platform accountability can strengthen collective defenses against visual misinformation (Brookings Institution, 2023).

8. Commitment to Corrections and Accountability: Policies must include clear procedures for handling errors, such as the inadvertent publication of AI-manipulated content. Swift retraction, transparent correction, and apologies are crucial for maintaining credibility (Jones et al., 2023). Accountability mechanisms reinforce ethical standards and demonstrate a commitment to truthfulness even when mistakes occur (Liu et al., 2023).

Conclusion

The integration of AI-generated imagery into the media landscape represents a critical juncture for journalism, testing its foundational ethical commitments. This technology introduces unprecedented capabilities for visual manipulation, posing significant threats through misinformation, challenging journalistic integrity, and complicating audience trust (Jones et al., 2023; Bendel, 2023). The potential for deepfakes and synthetic visuals to erode shared reality and exploit public credulity necessitates a robust, proactive response from the journalism community.

Existing scholarship and ongoing debates reveal a complex interplay between technological potential and ethical peril. While AI offers tools that could potentially enhance certain aspects of media production (Liu et al., 2023), its capacity for deception and bias demands stringent safeguards. The core journalistic values of truth, accuracy, transparency, fairness, and accountability remain the essential anchors in navigating this new terrain (Demmar & Neff, 2023). The challenge is not merely technological but deeply ethical, requiring a reaffirmation and adaptation of these principles to the context of synthetic media.

Patterns in the current discourse highlight a growing consensus on the need for strict verification, clear labeling, and restrictions on using AI visuals as factual reportage (Jones et al., 2023). However, gaps remain, particularly in developing universally understood labeling conventions, ensuring equitable application of standards across diverse global contexts (Sarhan & Hegelich, 2023), and adequately addressing the biases embedded within AI systems (Bendel, 2023). Furthermore, the rapid evolution of AI necessitates continuous adaptation of ethical guidelines and ongoing media literacy efforts for both journalists and the public (Arguedas & Simon, 2023).

A key ongoing challenge will be developing global standards for labeling and authentication that keep pace with increasingly sophisticated generation techniques. Moving forward, a multi-pronged approach involving rigorous internal policies, industry collaboration, technological solutions for authentication, and engagement with regulatory frameworks is essential. Journalists and news organizations must prioritize transparency, clearly communicating their standards and practices regarding AI to their audiences. By confronting the challenges of AI-generated imagery with ethical diligence and a steadfast commitment to truth, journalism can strive to maintain its credibility and fulfill its vital role in informing the public, even as the nature of visual information undergoes a profound transformation.

References

Amanbay, M. A. (2023). The ethics of AI-generated art: To what extent AI is a threat to artists? [Unpublished manuscript]. Nazarbayev Intellectual School of Physics and Mathematics in Aktobe.

Arguedas, A. R., & Simon, F. M. (2023). Automating democracy: Generative AI, journalism, and the future of democracy. University of Oxford.

Bauder, D. (2023, August 16). AP, other news organizations develop standards for use of artificial intelligence in newsrooms. Associated Press News.

Bendel, O. (2023). Image synthesis from an ethical perspective. AI & Society. Advance online publication.

Brookings Institution. (2023). Watch out for false claims of deepfakes, and actual deepfakes, this election year (D. S. Schiff & K. J. Schiff, Authors).

Demmar, K., & Neff, T. (2023). Generative AI in journalism education: Mapping the state of an emerging space of concerns, opportunities, and strategies. Journalism Education, 12(1), 47-58.

Eastwood, B. (2023, November 29). How should AI-generated content be labeled? MIT Sloan School of Management.

Habgood-Coote, J. (2023, June 9). What the deep history of deepfakes tells us about trust in images. LSE Impact Blog.

Ho, S. C. Y. (2023). From development to dissemination: Social and ethical issues with text-to-image AI-generated art. Canadian AI 2023.

Jones, B., Jones, R., & Luger, E. (2023). Generative AI & journalism: A rapid risk-based review. University of Edinburgh & BBC R&D.

Liu, J., Niu, Y., Jia, Z., & Wang, R. (2023). Assessing the ethical implications of artificial intelligence integration in media production and its impact on the creative industry. MEDAAD, 2023, 33-39.

Media, Entertainment & Arts Alliance. (2019). Journalist Code of Ethics.

National Press Photographers Association. (n.d.). Code of Ethics for Visual Journalists.

Palmer, B. (2023). Code of Ethics [Personal website declaration].

Paik, J., Cesare, A., & Betke, M. (2023). The Affective Nature of AI-Generated News Images: Impact on Audience Emotions and Perceptions (ACII 2023 Workshop paper).

Sarhan, H., & Hegelich, S. (2023). Understanding and evaluating harms of AI-generated image captions in political images. Frontiers in Political Science, 5.

Vyas, B. (2022). Ethical implications of generative AI in art and the media. International Journal for Multidisciplinary Research (IJFMR), 4(4), 1-11.