

Truth in Pixels: Generative AI, Synthetic Footage, and the Ethics of Authenticity in Documentary Filmmaking

Vedant Nitin Holkar¹

Bachelor of Applied Arts and Crafts, Final Year Student, Dr.
D. Y. Patil School of Design, Pune, India,

Siddhant Kumar Wadmare²

Vice Dean, Dr. D. Y. Patil School of Design,
Pune, India.

Abstract - Documentary filmmaking has long derived its moral authority from a foundational promise: that the camera bears witness to reality. This indexical contract — the implicit agreement between filmmaker and viewer that what appears on screen corresponds to something that actually occurred — is now under unprecedented strain. The emergence of generative artificial intelligence (AI) tools capable of producing photorealistic synthetic video, fabricating archival imagery, cloning voices, and digitally resurrecting deceased figures has introduced a profound epistemological challenge to non-fiction cinema. This paper critically examines the ethical dimensions of deploying AI-generated footage within documentary filmmaking, focusing on three core tensions: authenticity versus accessibility, transparency versus narrative immersion, and consent and the simulation of real persons. Drawing on documentary theory, media ethics scholarship, and emerging industry guidelines — including those of the Archival Producers Alliance (APA, 2024) and the British Broadcasting Corporation (BBC, 2024) — the paper proposes a tiered ethical framework for evaluating the permissibility of synthetic media in non-fiction contexts. The paper argues that generative AI does not inherently violate documentary ethics, but its deployment without robust transparency mechanisms, meaningful informed consent, and institutional accountability constitutes a serious breach of the genre's truth claim. The paper concludes with a call for enforceable international standards and audience literacy initiatives to safeguard documentary's epistemic integrity in the generative AI era.

Keywords: generative AI, documentary ethics, synthetic media, deepfakes, authenticity, non-fiction filmmaking, media transparency

1. INTRODUCTION

In 1922, Robert Flaherty staged scenes of Inuit life for *Nanook of the North*, a film widely regarded as the founding text of documentary cinema. The ethical controversy this staging provoked — and continues to provoke — speaks to a tension that has haunted non-fiction filmmaking since its inception: the gap between representing reality and constructing it. A century later, that tension has assumed a technological dimension of unprecedented complexity. Generative artificial intelligence now enables filmmakers, editors, and bad actors alike to synthesise footage of events that never occurred, voices that were never spoken, and faces that were never filmed. The implications extend far beyond aesthetics. Documentary film occupies a unique epistemic position in public culture. Unlike fiction, it makes explicit or implicit truth claims — asserting, or at minimum implying, that its images correspond to historical or contemporaneous reality (Nichols, 2010). Audiences approach documentaries with a fundamentally

different interpretive frame than they bring to drama or animation. It is this frame - this trust - that generative AI now threatens to destabilise.

Recent developments have accelerated urgency around this issue. In 2024, the Archival Producers Alliance released the first industry-wide guidelines for generative AI use in documentary filmmaking (APA, 2024). Major broadcasters including the BBC and PBS have issued internal policies governing AI-generated content (BBC, 2024). Meanwhile, films such as *Eno* (Hustwit, 2024) and *About a Hero* (Winiewicz, 2024) have used generative AI tools in ways that prompted significant critical debate. Academic scholarship has begun to address the intersection of deepfake technology and documentary ethics (Lees, 2025), but a comprehensive ethical framework for this domain remains elusive.

This paper aims to contribute to that framework. The central research question is: Does the use of AI-generated footage in documentary filmmaking constitute an ethical breach, and under what conditions — if any — can it be justified? The paper proceeds as follows. Section 2 surveys the relevant literature across documentary theory, media ethics, and AI scholarship. Section 3 maps the current landscape of AI tools used in documentary production. Section 4 analyses three core ethical tensions that AI-generated content creates. Section 5 reviews existing industry and regulatory frameworks. Section 6 proposes a tiered ethical model for evaluating synthetic media in non-fiction contexts. Section 7 concludes with reflections on the future of documentary's truth claim.

2. LITERATURE REVIEW

2.1. Documentary Theory and the Truth Claim

The theoretical foundation of this paper rests on Bill Nichols' influential taxonomy of documentary modes and his sustained engagement with documentary's relationship to reality. For Nichols (2010), documentary is defined not by formal properties but by the institutional and discursive context in which it circulates — a context that carries an implicit claim to represent the historical world. This claim is what Nichols calls the documentary's "voice": its assertive posture toward reality, which distinguishes it from fiction even when both employ similar techniques. Philip Rosen (2001) grounds this truth claim in the concept of indexicality, derived from C.S. Peirce's semiotics. Indexical signs — footprints, fingerprints, photographs — bear a causal or physical connection to their referents. The photographic image, in this view, is not merely a representation of reality but a trace of it: light that actually bounced off an actual subject inscribed onto a physical substrate. This indexical status has historically underpinned

documentary's epistemic authority. The camera does not just show us the world; it has been touched by it.

The rise of digital image-making has already complicated this indexical contract (Manovich, 2001; Rodowick, 2007). Digital images are fundamentally numerical rather than photochemical: they can be generated, modified, and assembled without any necessary connection to a profilmic reality. Scholars such as D.N. Rodowick (2007) argue that digital cinema represents a fundamental rupture with the indexical tradition, transforming the image from a trace of reality into a simulation of it. Generative AI intensifies this rupture to a qualitative degree. Where earlier digital manipulation required effortful post-production intervention, large language models and diffusion-based image generators can produce photorealistic imagery from text prompts alone.

2.2. Deepfakes and Media Ethics

Scholarship on deepfakes — synthetic media produced through deep learning techniques, typically involving the manipulation of human faces and voices — has grown substantially since the term entered public discourse around 2017 (Chesney & Citron, 2019). Chesney and Citron's foundational work identifies three primary harms of deepfake technology: facilitating non-consensual intimate imagery, enabling political disinformation, and eroding epistemic trust by creating a general atmosphere of uncertainty about the authenticity of all media — what they term the "liar's dividend."

In the documentary context, Lees (2025) has examined how filmmakers have been at the vanguard of adopting deepfake techniques, particularly for purposes of altering the face or voice of interview subjects — whether to protect identity, reconstruct deceased speakers, or correct technical problems in original footage. Lees argues that this adoption creates a constitutive tension with documentary's authenticity principle, but does not necessarily negate it: the ethical valence of the technique depends substantially on context, consent, and disclosure.

The question of consent is particularly underexplored in existing scholarship. While bioethics and research ethics have developed robust frameworks for informed consent in contexts involving human subjects, media ethics has been slower to theorise consent in relation to the representation of real persons in documentary. This gap is especially significant when AI tools can reconstruct or simulate the likeness of individuals — including deceased persons who cannot consent — with increasing verisimilitude.

2.3. AI and Post-Production Automation

Beyond deepfakes narrowly construed, a broader literature addresses the integration of machine learning into post-production workflows. Scholars and practitioners have examined AI-driven tools for automated editing (Hausmann et al., 2019), music synchronisation, colour grading, and subtitle generation. While this literature tends toward the technical and evaluative rather than the ethical, it provides essential context for understanding the range of AI interventions now available to documentary makers.

Of particular relevance is the use of AI for generating or augmenting archival material. Archival footage and photography are foundational to historical documentary, yet archives are radically uneven: they over-represent the powerful, the Western, and the photogenic, while systematically under-documenting subaltern experience, pre-photographic history, and events that occurred beyond the camera's reach. Generative AI creates, for the first time, a technical possibility of filling these gaps — of

depicting visually what was never filmed. The ethical dimensions of this possibility are only beginning to be theorised.

3. The Landscape of AI Tools in Contemporary Documentary

The AI tools now available to documentary filmmakers span a wide spectrum of intervention, from subtle enhancement to wholesale fabrication. It is important to map this spectrum clearly before proceeding to ethical analysis, since the ethical stakes vary considerably across different types and degrees of AI involvement. At one end of the spectrum lie tools that augment or improve existing footage without generating new visual content: AI-powered noise reduction, image stabilisation, upscaling of low-resolution archival material, and automatic colour matching. These applications are largely uncontroversial and have been broadly adopted across the industry. They enhance what already exists rather than constructing what does not.

More ethically complex are tools that modify existing footage in ways that alter meaning. Voice cloning technology can reconstruct the voice of a subject from limited audio samples, enabling filmmakers to create dialogue or narration that the subject never actually recorded. Facial re-animation tools — a subset of deepfake technology — can animate still photographs or apply one person's facial movements to another's recorded face. These techniques have been used to give voice to historical figures in documentary contexts: the AI-generated reconstruction of Werner Herzog in *About a Hero* (Winiewicz, 2024) is a prominent recent example (International Documentary Association, 2024).

At the far end of the spectrum lie fully generative tools: diffusion models such as Stable Diffusion, Midjourney, and Sora that can produce video or still imagery from text descriptions alone. The use of such tools to create synthetic "archival" material — images purporting to depict historical events or periods — represents perhaps the most serious ethical challenge for documentary. As the APA documented in its 2024 guidelines, filmmakers have already used text-to-image generation to produce imagery depicting historical eras with sparse photographic records, imagery that is visually indistinguishable from genuine archival photographs (APA, 2024).

It should also be noted that AI is increasingly embedded in production tools that filmmakers may use without fully registering as AI: smart auto-cut features in editing software, AI-driven transcription and translation services, and automated subtitle generation. The boundaries between AI and non-AI production are becoming increasingly porous, complicating any framework that attempts to regulate a discrete category of "AI-generated content."

4. CORE ETHICAL TENSIONS

4.1. Authenticity Versus Accessibility

The first and most fundamental tension concerns the relationship between documentary's truth claim and the creative possibilities that AI opens up. Documentary's epistemic authority rests on authenticity — the assurance that its images correspond to a historical reality that the filmmaker has witnessed or recovered. Yet this commitment to authenticity has always been in tension with the practical and material constraints of filmmaking: the footage that was not shot, the archive that does not exist, the event that occurred before the camera was invented.

Generative AI offers a potential resolution to this tension. By enabling filmmakers to visualise what was never filmed, it promises greater accessibility to marginalised histories, under-documented communities, and pre-photographic periods. A

documentary about the Middle Passage, for instance, could use AI-generated imagery to visualise the interior of a slave ship — an experience that was never, and could never have been, captured on film. The argument for such use invokes documentary's pedagogical and social justice functions: if the purpose of documentary is to illuminate reality for audiences, AI-generated images that accurately represent historical reality (even if not indexically derived from it) serve that purpose.

The counter-argument is that such use fundamentally alters the epistemic status of the documentary image in ways that audiences may not register and cannot evaluate. When viewers see imagery in a documentary, they make involuntary indexical inferences: they assume that someone, somewhere, pointed a camera at a corresponding reality. Violating this assumption without clear disclosure is not merely a formal transgression but a cognitive deception — it shapes the audience's relationship to historical knowledge in ways they have not consented to. Moreover, the apparently "accurate" representation of historical reality through AI generation is itself epistemically problematic: it presents a particular imaginative rendering as if it were documented fact, potentially supplanting more accurate or more capacious historical understandings.

4.2. Transparency Versus Narrative Immersion

Even if the ethical justification for AI-generated content can be established in particular cases, a second tension arises around the obligation of transparency. The principle that audiences should be informed when they are viewing synthetic rather than indexical imagery commands near-universal assent in industry discussions (APA, 2024; BBC, 2024). The disagreement concerns how this transparency should be operationalised in practice.

The most straightforward approach — explicit on-screen labelling of AI-generated content — faces significant practical objections. As APA guidelines acknowledge, intrusive watermarks or labels can disrupt narrative immersion and undermine the emotional and aesthetic impact that gives documentary its communicative power. A film documenting atrocity, for instance, may lose its affective force if viewer engagement is constantly interrupted by disclosure notices (APA, 2024).

The APA's proposed solution — less obtrusive disclosure conventions, such as distinctive framing, colour grading, or aspect ratio changes for AI-generated sequences — represents a pragmatic compromise. Yet this approach places considerable demands on audience media literacy: it assumes that viewers will recognise and correctly interpret these conventions, an assumption that cannot be taken for granted across diverse audience populations. PBS has gone further, requiring disclosure in closing credits, top-of-show notices, and supplemental online materials — an approach that distributes transparency across multiple platforms and access points (Nieman Lab, 2024).

The transparency dilemma is further complicated by the temporal dynamics of disclosure: pre-release disclosure, in-film disclosure, post-film disclosure, and platform-level metadata all operate differently in terms of their impact on audience experience and epistemic position. A viewer who learns that a documentary contains AI-generated content before watching it will approach the film differently from one who discovers this fact in the closing credits. Neither mode of disclosure is self-evidently superior; both involve trade-offs between experiential authenticity and informational completeness.

4.3. Consent and the Simulation of Real Persons

The third and arguably most legally and ethically fraught tension

concerns the use of AI to simulate, reconstruct, or animate real persons — whether living or deceased. Voice cloning, facial re-animation, and AI-generated "performances" all raise profound questions about consent, identity, and the right to control one's own likeness and voice.

For living subjects, the consent framework is relatively clear in principle if not always in practice: using AI to simulate a real person's voice or face without their knowledge and agreement is a violation of their autonomy and, in many jurisdictions, their legal rights. The complications arise when subjects are deceased, when their estates claim control over their likeness, or when documentary makers invoke journalistic or public interest justifications for simulation without consent.

The case of deceased historical figures is particularly complex. A documentary that uses AI to animate Abraham Lincoln delivering a speech he actually gave raises different ethical questions than one that has him deliver a speech he never gave; yet both deploy AI simulation of a real person to generate imagery that never existed. The former may be understood as a form of historical recreation; the latter shades into fabrication. The distinction matters, but the visual appearance of both on screen may be identical, and audience members are unlikely to possess the contextual knowledge needed to discriminate between them.

The simulation of living subjects without consent introduces additional harms. As Chesney and Citron (2019) observe, deepfake technology can be weaponised to damage reputations, manufacture evidence, and violate the bodily and psychological integrity of individuals. Even in benign documentary contexts, AI simulation of a living person implies a claim about what that person said, did, or experienced that may be inaccurate or unwelcome. The International Documentary Association has identified this as an area of particular concern in its 2025 survey of filmmaker perspectives (IDA, 2025).

5. EMERGING REGULATORY AND INDUSTRY FRAMEWORKS

The ethical challenges outlined above have begun to generate regulatory and industry responses, though these remain fragmented and largely voluntary. This section surveys the principal frameworks currently in operation and identifies their strengths and limitations.

The most comprehensive industry-specific guidance to date has been produced by the Archival Producers Alliance. Released at the Camden International Film Festival in September 2024, the APA Best Practices constitute the first industry-wide guidelines for generative AI use in documentary. Their central principle is "Outward Transparency": the APA insists that audiences must never be confused about what is generative AI and what is not (APA, 2024). The guidelines draw important distinctions between AI-generated content used for artistic interpretation (permissible with disclosure) and synthetic archival material that could be mistaken for genuine historical documentation (prohibited or subject to stringent constraints).

The BBC's 2024 AI guidelines, while not documentary-specific, establish broad principles applicable to non-fiction content: AI-generated content must be clearly identified to audiences; it must not undermine audience trust; and all AI use must be subject to effective human oversight (BBC, 2024). PBS has adopted a more detailed disclosure toolkit specifying multiple modes of audience notification for AI-generated content (Nieman Lab, 2024).

At the regulatory level, the European Union's AI Act (2024) establishes requirements for the labelling of deepfakes and AI-generated content, though its application to artistic and

documentary contexts involves significant interpretive complexity. The Act's provisions for "high-risk" AI systems — those deployed in contexts with significant potential for harm — do not straightforwardly encompass documentary filmmaking, and the Act's treatment of artistic expression suggests a degree of carve-out for creative applications. In the United States, no comparable federal legislation exists, though several states have enacted deepfake-specific laws with implications for documentary production.

The limitations of existing frameworks are significant. They are largely voluntary, industry-specific, and jurisdiction-dependent. They assume a degree of audience media literacy that may not exist. They do not adequately address the cumulative epistemological effects of widespread synthetic media use across the documentary ecosystem — the gradual erosion of audience trust in non-fiction imagery that may result not from any individual act of deception but from the aggregate uncertainty generated by AI-saturated media environments.

6. A Proposed Tiered Ethical Framework

Drawing on the theoretical and empirical analysis presented in the preceding sections, this paper proposes a tiered ethical framework for evaluating the permissibility of AI-generated content in documentary filmmaking. The framework is organised around three tiers, defined by the nature and degree of ethical risk involved, and the corresponding conditions that must be satisfied for use to be justified.

6.1. Tier 1 — Permissible with Disclosure

The first tier encompasses AI uses that enhance or augment existing footage without generating substantively new visual or auditory content, and AI uses that visualise historically or empirically grounded information for which no documentary footage exists. Examples include: AI upscaling of archival video, AI noise reduction, AI colour correction, and AI-generated visual representations of documented historical events or settings for which no photographic record exists.

The justification for permissibility in this tier rests on the argument that the epistemic function of documentary — the communication of historical or contemporaneous reality to audiences — is served rather than subverted by these uses, provided that audiences are clearly informed. The condition of disclosure is non-negotiable: uses in this tier must be identified to audiences through appropriate means, whether in-film, in credits, or in supplemental materials, and the disclosure must be sufficiently clear to enable informed audience interpretation.

6.2. Tier 2 — Permissible with Consent and Disclosure

The second tier encompasses AI uses that reconstruct, simulate, or animate identifiable real persons — including voice cloning, facial re-animation, and AI-generated performances. These uses carry a higher degree of ethical risk because they implicate the autonomy and identity of specific individuals and are more likely to generate false impressions about what those individuals actually said or did.

Uses in this tier are permissible only where meaningful informed consent has been obtained from the subject (or, in the case of deceased persons, from an appropriate representative and subject to ethical review), where the purpose is clearly in the public interest, and where robust disclosure mechanisms ensure that audiences understand they are viewing simulated rather than documentary footage. The consent obtained must be specific: general consent to participate in a documentary does

not constitute consent to AI simulation of one's voice or likeness.

6.3. Tier 3 — Prohibited

The third tier encompasses AI uses that fabricate events, statements, or actions by real persons in ways that are presented — or are likely to be perceived — as factual documentation. The creation of synthetic footage depicting a real person committing an act they did not commit, making a statement they did not make, or being present at an event they did not attend falls within this tier, regardless of the filmmaker's intentions. Similarly prohibited is the use of AI-generated imagery to simulate historical events or conditions in ways that are not grounded in documented historical evidence and are likely to be mistaken for genuine archival material.

The prohibition in this tier is absolute: no degree of disclosure, consent, or public interest justification is sufficient to render these uses ethically permissible within the genre of documentary. They constitute not merely formal transgressions but fundamental violations of documentary's constitutive truth claim and of the audience's epistemic rights.

7. CONCLUSION

The emergence of generative AI as a tool of documentary production has not created the ethical challenges this paper has examined — it has intensified and accelerated them. Documentary filmmaking has always navigated the tension between representing reality and constructing it, between the indexical authority of the camera and the selective, perspectival nature of all representation. What AI has done is to make the most extreme forms of construction — the wholesale fabrication of events, voices, and persons — technically trivial and visually indistinguishable from genuine documentation.

The framework proposed in this paper is an attempt to provide principled guidance for navigating this landscape. Its central argument is that generative AI does not inherently violate documentary ethics: there are uses of synthetic media that serve documentary's truth-telling function, provided they are bounded by robust transparency, meaningful consent, and institutional accountability. What is prohibited — and what existing frameworks have sometimes failed to prohibit clearly enough — is the use of AI to fabricate historical reality, to simulate real persons without consent, and to deceive audiences about the epistemic status of what they are watching.

The implementation of such a framework requires action at multiple levels. Industry bodies must move from voluntary guidelines toward enforceable standards, with mechanisms for accountability when those standards are breached. Regulatory frameworks must be adapted to address the specific conditions of documentary production, with appropriate carve-outs for artistic expression that do not become loopholes for deception. And filmmakers, archivists, and educators must invest in the audience literacy that disclosure alone cannot provide — the critical capacity to engage thoughtfully with the epistemological complexity of an image environment in which the real and the synthetic are increasingly intertwined.

The stakes are not merely aesthetic. Documentary film has been instrumental in shaping public understanding of history, politics, and human experience. The erosion of its truth claim — whether through individual acts of fabrication or through the cumulative

uncertainty generated by an AI-saturated media environment — is a loss not just for the genre but for public epistemology. Protecting that truth claim in the age of generative AI is among the most urgent tasks facing the documentary community.

8. REFERENCES

- (1) Archival Producers Alliance (APA). (2024). Best practices for the use of generative AI in documentaries. Archival Producers Alliance.
- (2) British Broadcasting Corporation (BBC). (2024). Guidance: The use of artificial intelligence. BBC Editorial Guidelines. <https://www.bbc.co.uk/editorialguidelines/guidance/use-of-artificial-intelligence>
- (3) Chesney, R., & Citron, D. K. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. *California Law Review*, 107(6), 1753–1820.
- (4) European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.
- (5) Flaherty, R. (Director). (1922). *Nanook of the North* [Film]. Pathé Exchange.
- (6) Harari, Y. N. (2024). *Nexus: A brief history of information networks from the Stone Age to AI*. Signal Press.
- (7) Haussmann, S., Paudel, B., & Bernstein, A. (2019). Automated video summarisation for documentary films. *Proceedings of the 2019 ACM International Conference on Multimedia*, 2117–2125.
- (8) Hustwit, G. (Director). (2024). *Eno* [Film]. Oscilloscope Laboratories.
- (9) International Documentary Association (IDA). (2025, April). The synthesis: All bets are off for AI and documentary — it is time for a reset. *Documentary Magazine*. <https://www.documentary.org/column/synthesis-all-bets-are-ai-and-documentary-its-time-reset>
- (10) Lees, D. (2025). Artificial intelligence in documentary filmmaking: The ethics of deepfakes. In A. Piotrowska (Ed.), *The ethics of documentary film* (pp. 296–318). Edinburgh University Press.
- (11) Manovich, L. (2001). *The language of new media*. MIT Press.
- (12) Nichols, B. (2010). *Introduction to documentary* (2nd ed.). Indiana University Press.
- (13) Nieman Journalism Lab. (2024, September 12). Documentary filmmakers publish new AI ethics guidelines: Are news broadcasters next? <https://www.niemanlab.org/2024/09/documentary-filmmaker-alliance-sets-ai-ethics-guidelines-a-lesson-for-news-broadcasters/>
- (14) Peirce, C. S. (1931–1958). *Collected papers of Charles Sanders Peirce* (Vols. 1–8, C. Hartshorne, P. Weiss, & A. W. Burks, Eds.). Harvard University Press.
- (15) Rodowick, D. N. (2007). *The virtual life of film*. Harvard University Press.
- (16) Rosen, P. (2001). *Change mummified: Cinema, historicity, theory*. University of Minnesota Press.
- (17) Winiewicz, P. (Director). (2024). *About a hero* [Film]. Presented at IDFA.