Routledge
Taylor & Francis Group

Information, Communication & Society

Synthetic humanitarianism and algorithmic witnessing: generative AI and the politics of humanitarian representation

Submission ID	260489592
Article Type	Research Article
Keywords	synthetic humanitarianism, algorithmic witnessing, generative AI, humanitarian communication, technocolonialism, decolonial theory
Authors	Marco Scalvini

Abstract

Humanitarian organisations have begun deploying generative AI imagery in humanitarian campaigns, a practice this article terms synthetic humanitarianism. The article asks what conceptual framework is required to analyse the ethical and communicative implications of this practice. Drawing on witnessing theory, decolonial critiques of humanitarian visibility, and technocolonialism, I argue that synthetic humanitarianism is not a novel disruption but an intensification of extractive and colonial logics that have structured humanitarian communication since its institutional emergence. I develop the concept of algorithmic witnessing to name the communicative condition in which generative AI systems produce the relations of testimony and moral address without any act of witnessing having taken place and without the participation of those represented. The framework specifies four dimensions, testimonial fabrication, epistemic extraction, delanguaging, and stereoscopic automation, and I apply it to thirteen documented cases of generative AI image use in humanitarian communications between 2022 and 2025. The analysis demonstrates that these dimensions form a mutually reinforcing system in which the elimination of situated meaning-making enables fabricated testimony and the computational encoding of colonial visual archives produces automated stereotype. I argue that the most significant consequence of synthetic humanitarianism is political rather than epistemological. The displacement of situated meaning-making by algorithmic production eliminates the last structural point at which affected communities could intervene in their own representation.

Keywords: synthetic humanitarianism, algorithmic witnessing, generative AI, humanitarian communication, technocolonialism, decolonial theory

Synthetic humanitarianism and algorithmic witnessing:

generative AI and the politics of humanitarian representation

Marco Scalvini

Introduction

Humanitarian communication has long depended on the photographic image as its primary instrument of moral address. This dependence rests on a particular understanding of witnessing. Witnessing is a fundamental communicative "for securing truth from the facts of our sensitivity to pain and our inevitable death, increases the stakes of our thinking about media events"(Peters, 2001, p. 709). In humanitarian contexts, audiences do not encounter suffering directly. They witness someone else's witnessing, mediated through what Silverstone (2007) calls the "mediapolis," the mediated public space in which photography, television, and digital platforms frame how distant events are encountered and interpreted. The photograph, in this tradition, does not illustrate a crisis. It functions as testimony, placing a moral demand on the spectator to recognise and respond.

Since late 2022, a fundamentally different mode of image production has entered this communicative architecture. As the corpus analysed in this article documents, humanitarian organisations and campaign actors have begun deploying AI-generated imagery in fundraising appeals and social media campaigns. These images depict photorealistic scenes that never occurred, featuring persons who do not exist, yet they are designed to perform the same testimonial function as documentary photography. I define this emergent practice as synthetic humanitarianism, the strategic deployment of computationally generated imagery to perform humanitarian communication functions historically reserved for media that document real events

and real persons. Where humanitarian communication has always presented mediated witnessing, synthetic humanitarianism presents fabricated witnessing. The spectator is addressed as though testimony has taken place, but no communicative act of witnessing preceded the image.

Existing scholarship has not accounted for this shift. The dominant critical tradition has traced a trajectory from "compassion fatigue" (Moeller, 1999) through the spectacle of suffering (Chouliaraki, 2006) to the post-humanitarian condition, in which solidarity operates through what Chouliaraki (2013, p. 55) describes as a "low-intensity form of engagement" and practices of playful consumerism. More recent scholarship has demonstrated that witnessing is shaped by mediatisation, that audiences negotiate their moral positioning toward distant events through culturally and historically specific practices of media reception (Kyriakidou, 2015, pp. 215–217). Each of these contributions, however, presupposes that the image under analysis retains a referent in lived experience. However contested or instrumentalised the representation, the visual record originates in an encounter between a recording technology and an actual state of affairs. The conceptual vocabulary available within this tradition, including the "analytics of mediation" (Chouliaraki, 2006, p. 70) and the "politics of pity" (Boltanski, 1999, p. 11), was developed to analyse mediated witnessing, not its synthetic production.

Synthetic humanitarianism eliminates the testimonial ground on which this vocabulary depends. The image performs the communicative function of testimony. It addresses spectators as witnesses and demands moral response. But no act of witnessing preceded it because there was nothing to witness. This is not a further weakening of the "testimonial relation that constitutes witnessing, but its evisceration" (Richardson, 2024, p12). The consequences extend beyond epistemology. Moral response depends on a prior determination of what Butler (2009, p.

1) terms “grievability,” on perceptual frames that regulate which lives are recognised as living and which are effaced. In documentary humanitarian communication, however contested the framing, the image refers to a life that exists and could in principle be recognised as grievable. Synthetic humanitarianism inverts this relation. The algorithm generates a grievable subject who has no ontological existence. Grievability is manufactured rather than recognised, and those on whose behalf it is manufactured remain structurally unable to contest how, or whether, they are represented. Critical data studies have theorised how what Couldry and Mejias (2019, p. 5) term “data colonialism” extends colonial logics of extraction into the domain of everyday life, but have not addressed the specific conditions under which fabricated visibility acquires moral force and humanitarian authority.

This article asks what conceptual framework is required to analyse the ethical and communicative implications of deploying generative AI imagery in humanitarian communication. I develop a framework organised around the notion of algorithmic witnessing[1] and apply it to thirteen documented cases of generative AI image use in humanitarian communications between 2022 and 2025, following the method of critical conceptual analysis in which illustrative instances are examined to develop and refine theoretical propositions.

The article proceeds as follows. The second section develops the theoretical framework through three cumulative movements, arriving at the concept of algorithmic witnessing. The third section specifies the methodology. The fourth section applies the framework to the case corpus. The fifth section examines the structural contradictions that the framework exposes. I

[1]The phrase “witnessing algorithms” has been used in nonhuman witnessing scholarship to examine the witnessing performed by and of algorithms (Richardson, 2024, Chapter 2, p. 80). I use the term in a distinct sense to name the communicative condition in which the relations of humanitarian testimony are computationally produced without any act of witnessing having taken place.

conclude by arguing that synthetic humanitarianism demands conceptual tools adequate to a communicative environment in which the suffering body is produced rather than documented.

Theoretical framework

The preceding section identified the epistemological rupture that synthetic humanitarianism introduces into humanitarian witnessing. That rupture, however, is not self-explanatory. To treat it solely as a novel problem arising from a novel technology would be to mislocate its origins. The conditions enabling synthetic humanitarianism are embedded in the epistemic and representational structures through which humanitarian communication has operated since its nineteenth-century emergence. The testimonial model of witnessing, the regimes of visibility through which humanitarian organisations represent distant suffering, and the infrastructural technologies through which displaced communities are governed all contain colonial logics that the dominant literature has left largely unexamined. The framework that follows develops this argument through three cumulative movements, arriving at the concept of algorithmic witnessing as the conceptual intervention required to analyse this condition in both its communicative and colonial dimensions.

The conditions of testimonial witnessing and what they conceal

The communicative structure of witnessing involves a distinction that carries significant consequences for humanitarian practice. Two modalities, what Peters (2001, p. 709) calls the "two faces" of witnessing, are often conflated but must be separated analytically (Frosh & Pinchevski, 2009, pp. 1–2). The first is the "passive one of seeing," the fact of having been present when an event occurred. The second is the "active one of saying," an active discursive performance in which the person who was present translates that experience into communicable form for an audience that was absent. The gap between these two modalities is irreducible.

Testimony always involves transformation at what Peters (2001, p. 710) calls the "difficult juncture between experience and discourse." In humanitarian communication, this gap is compounded by institutional mediation. What audiences engage in is not witnessing in the strict sense but the mediation of distant suffering, a process shaped by the communicative decisions of intervening institutions (Ong, 2014). The spectator who encounters a campaign image witnesses neither the event itself nor the direct testimony of a person who was present, but an institutionally produced artefact that claims testimonial authority on behalf of absent others. For this mediated testimony to function, several conditions must hold simultaneously. A real event must have occurred. A presence, human or technological, must have registered it. And the communicative artefact must maintain a traceable relation to that originating encounter, however attenuated.

The normative demands of witnessing extend beyond this evidentiary structure into the ethical disposition of the spectator. Cosmopolitan accounts of witnessing insist that genuine engagement with distant suffering requires an imaginative effort to see "from the standpoint of those who are absent" (Robertson & Schaetz, 2021, p. 324). The spectator must recognise that the suffering other inhabits a context with its own coherence and its own modes of understanding, and that witnessing carries an obligation to engage with that context rather than to consume its visible surface. Photographs of suffering do not automatically produce ethical response; the relation between image and moral obligation is neither natural nor guaranteed (Sontag, 2003, p. 7). This standard has been consistently undercut in humanitarian communication practice, where the predominant visual strategies have subordinated the other's context to the spectator's emotional and moral needs (Ongenaert & Soler, 2024; Chouliaraki,

2006; Seu & Orgad, 2014). The failure is not incidental but symptomatic of a deeper structural problem in the testimonial relation itself.

That structural problem becomes visible when witnessing is analysed as a relation of power. The analysis of media as governance reveals communication itself as a mode of control over the other, in which the capacity to frame, narrate, and circulate representations of trauma operates as a form of authority (Heller, 2014, p. 12). Trauma, in this analysis, can be inverted into what Heller (2014, pp. 86–102) calls a "pedagogical requirement," dictating how the traumatised must perform their suffering in order to receive recognition. Humanitarian organisations do not passively record the experiences of those they represent. They establish the conditions of appearance, the terms of legibility, and the frameworks of interpretation through which the represented subject becomes visible. The testimonial relation in humanitarian witnessing is therefore a relation in which the witnessing party holds epistemic and political authority over the witnessed. This asymmetry is not an accidental distortion of humanitarian communication but a structural feature that the dominant witnessing literature has treated as incidental rather than constitutive.

The full significance of this asymmetry emerges when it is read through what Mignolo (2009, p. 160) terms the "hubris of the zero point." The zero point names the unmarked epistemic position from which Western knowledge systems observe, classify, and evaluate while concealing their own geopolitical location (Mignolo, 2009, p. 161). The humanitarian witness occupies precisely this position. The spectator who encounters an image of suffering in the Global South is positioned as a moral agent whose concern is universal and whose gaze is ethically motivated, while the geopolitical coordinates of that gaze remain invisible. The distant other is made available for moral consumption from a position that presents itself as universal

concern but that operates as epistemic extraction. The testimonial model of witnessing, even in its most ethically sophisticated formulations, therefore contains a colonial structure it has not theorised as such. The zero point is the epistemic condition that makes humanitarian witnessing possible in its current form.

The extractive architecture of humanitarian visibility

If the zero point describes the epistemic position from which humanitarian witnessing operates, the question that follows is how this position materialises in the representational practices of humanitarian organisations. I turn now from the structure of the testimonial relation to the regime of visibility through which humanitarian actors determine who appears, in what form, and for whose benefit.

Humanitarian visibility operates as a selective regime that generates specific forms of appearance while producing disappearance. The history of humanitarian photography demonstrates that this selectivity is a structural feature of humanitarian visual practice since its emergence (Fehrenbach & Rodogno, 2015, pp. 1–3). International humanitarian campaigns construct figures of "ideal victims," overwhelmingly women and children rendered as helpless and passive, and circulate them as hypervisible in order to mobilise donor publics in the Global North (Fiddian-Qasmiyeh, 2019, pp. 28–32). This hypervisibility is structurally paired with a corresponding invisibility. The labour of local refugee communities, the mutual aid networks through which displaced people sustain one another, and the structural conditions that produced displacement fall outside the frame of institutional representation. The actual burden of care and survival is privatised, absorbed by communities whose work remains unseen (Fiddian-Qasmiyeh, 2019, pp. 30–34). Humanitarian organisations do not extend visibility to those who suffer. They

extract it from specific bodies and specific configurations of distress because those bodies are useful to the institutions that circulate them.

The representational strategies through which this extraction operates follow identifiable patterns. Whether through the shock effect of suffering crowds or the deliberate positivity of empowered beneficiaries, the refugee functions as raw material for institutional communicative objectives rather than as an agent within their own narrative (Ongenaert & Soler, 2024). In post-humanitarian communication, this displacement intensifies further (Chouliaraki, 2013, p. 2). The focus shifts to the donor's emotional experience, the charity's brand identity, and the celebrity who lends visibility to the cause. I read this trajectory not as a decline from an earlier ethical standard but as the progressive clarification of a representational economy in which humanitarian organisations have always produced and consumed the other's visibility according to institutional need.

The decolonial stakes of this economy become fully legible through what Mignolo (2021, p. 205) terms the "logic of the in-visible." The Western drive to render the world visible, knowable, and classifiable is itself a colonial operation (Mignolo, 2021, p. 215). Visibility on these terms requires the subject to submit to categories of legibility that serve the observer's epistemic and moral economy. What Fassin (2012, p. xi) terms "humanitarian reason," a generalised mode of government deploying the language of suffering and compassion, operates precisely through this economy. The suffering other must expose their misfortune to arouse sympathy and justify aid (Fassin, 2012, p. 193), and the power relations that structure this encounter remain concealed by the moral sentiment that authorises it. The represented subject must appear as a recognisable type, the starving child, the grateful beneficiary, the resilient survivor, or forfeit visibility altogether. Decoloniality, in this framework, requires recognising

the subject's right to remain in-visible to this extractive matrix, to exist on epistemic terms not reducible to Western frameworks of recognition (Mignolo, 2021, p. 205).

The concept of "decolonial pluriversality" (Mignolo & Walsh, 2018, p. 2) provides the horizon against which humanitarian communication's universalising logic can be evaluated. Humanitarian representation has functioned as a regime that forecloses pluriversal modes of self-representation. Affected communities do not determine the terms of their own appearance. They appear, if at all, in forms that the Western spectator can process within an existing moral vocabulary.

This representational foreclosure has a material and infrastructural dimension. The media infrastructures through which migration has been governed, from paper identity documents to contemporary digital databases, function as systems designed for the "efficient ordering, administrating and limiting of refugee bodies" (Seuferling & Leurs, 2021, p. 671). Through these infrastructural capacities, refugee subjects were reduced to administrable data points, Chouliaraki & Zaborowski (2017, p 616) discuss how refugees are often reduced to a "statistical percentage" or an "anonymous mass", stripped of political agency by humanitarian and state institutions alike. This reveals a continuity between the bureaucratic reduction of persons to administrative data and the computational reduction of persons to training data within generative AI systems.

I argue that humanitarian visibility has operated as an extractive regime since its institutional emergence. Synthetic humanitarianism does not introduce extraction into an otherwise ethical system. It intensifies and automates an extraction that was already operative.

Technocolonialism and the synthetic production of the other

The continuity between bureaucratic and algorithmic governance identified in the preceding section takes on a qualitatively new character with the emergence of generative AI in humanitarian communication. The conceptual resources for analysing this shift already exist in part. The concept of technocolonialism describes how the "convergence of digital developments with humanitarian structures and market forces" reinvigorates and reshapes colonial relationships of dependency between aid organisations and refugee communities (Madianou, 2019, p. 2). Humanitarian technologies, from biometric registration systems to data-sharing agreements between aid agencies and technology companies, do not function as neutral tools adopted for operational efficiency. They constitute an extractive infrastructure in which aid organisations and their technology partners harvest the data of displaced people, process it in the Global North, and monetise it through systems over which the people from whom it originates exercise no control (Madianou, 2019, pp. 4–8). I build on this framework and extend it to a domain the original analysis did not address. Generative AI imagery in humanitarian communication represents a further stage of technocolonial extraction, one in which the raw material is no longer biometric data or registration information but the accumulated visual record of suffering itself.

The extraction begins with the construction of training datasets. Technology companies scrape billions of images from the internet, incorporating vast quantities of photographs depicting vulnerable people, refugees, and communities in crisis, without informing the persons depicted or obtaining their consent (Crawford, 2021; Birhane & Prabhu, 2021). The extraction operates at the epistemic level. What is mined is not simply pixel data but the visible traces of lived experience that documentary photography recorded over decades of humanitarian practice. These traces are converted into computational capital, statistical patterns that Bender et al. (2021, p. 614) define as "stochastic parrots," generating outputs by statistical extrapolation rather than

by comprehension. The technocolonial logic identified in biometric systems (Madianou, 2019, p. 10) is reproduced here in intensified form. Where biometric registration extracts identity data from living refugees who must submit to the system in order to receive aid, generative AI training extracts representational data from the entire historical archive of humanitarian imagery, including images of persons who may be dead or displaced beyond traceability.

This extraction acquires its full significance when the extracted data generates the representations that replace the communities from which it was taken. Before generative AI, humanitarian imagery required some form of encounter with the real, however asymmetrical. Synthetic imagery eliminates this encounter entirely. An agency worker in a Western office types a text prompt and produces the desired image without any contact with those represented. I read this severance through the concept of "languaging" (Mignolo & Walsh, 2018, p. 171), the active, situated, and bodily practice of making meaning that is crucial for enacting a decolonial relational world. Synthetic image production constitutes a form of what I term delanguaging, the systematic removal of situated, embodied meaning-making from those whose lives are being represented. Delanguaging is not a failure of communication. It is the elimination of the conditions under which the other could communicate at all. The observation that technocolonial systems create dependency by making humanitarian subjects reliant on infrastructures they cannot control and "instruments of surveillance and governance over displaced populations" (Madianou, 2025, p. 129) finds its extreme expression here. The dependency is no longer on a data system the refugee must submit to in order to receive aid but on a representational system that no longer requires the refugee's presence, participation, or consent at any stage.

When delanguaging removes the capacity for situated meaning-making, what fills the representational void is the automated reproduction of colonial visual logic. Generative AI

models default to the visual tropes that dominate their training data, which in humanitarian representation means the iconography of helpless victims, passive refugees, and starving children (Manzo, 2008, pp. 632–635; Dogra, 2012). Empirical studies have confirmed a persistent "White default" in AI-generated imagery (Park, 2024, p. 38) and systematic racial bias that users themselves perceive as problematic (Messingschlager & Appel, 2025). Synthetic humanitarianism automates the production of these stereotypes at scale, encoding a Western imagination of crisis as the default visual grammar of humanitarian representation. The diverse and pluriversal realities of affected communities are excluded by the model's statistical architecture before any specific image is generated. This is not a technical flaw amenable to better training data but the structural consequence of a system built on colonial archives (Mignolo & Walsh, 2018, p. 200), operated from the zero point, and deployed in the absence of any situated representational process that could contest its outputs.

I argue, therefore, that synthetic humanitarianism is a technocolonial practice that operates through a closed extractive cycle. The suffering other is first extracted as training data, then replaced by their own algorithmic reproduction through a process of delanguaging that eliminates embodied meaning-making, and finally locked into automated colonial stereotypes that ban pluriversal self-representation. The cycle is closed because the communities whose documented suffering provided the raw material for generative AI are now represented by the systems that consumed them, and the representations produced owe nothing to their knowledge or their consent. This is the condition that the concept of algorithmic witnessing is designed to name.

Algorithmic witnessing as conceptual framework

The preceding analysis has demonstrated that the problem of synthetic humanitarianism cannot be adequately addressed by existing theoretical resources. The witnessing literature provides the communicative architecture but has not theorised its own colonial structure. The decolonial critique of humanitarian visibility exposes the extractive regime within which representations of suffering are produced and consumed but does not account for the synthetic fabrication of those representations. The technocolonial analysis of humanitarian technology identifies the extractive logic of digital systems but has not yet been extended to the generative production of imagery that replaces the communities from which its training data was extracted. Each body of scholarship illuminates a dimension of the problem. None addresses the condition as a whole. I develop the concept of algorithmic witnessing to address this gap.

I define algorithmic witnessing as the communicative condition in which the relations of testimony and moral address that constitute humanitarian witnessing are computationally produced without any act of witnessing having taken place and without the participation or situated knowledge of those represented. Algorithmic witnessing names not a technology but a reconfiguration of the witnessing relation. The spectator is still addressed as a witness. The image still performs the communicative function of testimony. But the testimonial ground has been eliminated. No event was recorded. No situated meaning-making process connected the representation to the lived reality of those depicted. The witnessing relation is fabricated from within, produced by a computational system that simulates the communicative form of testimony while severing every connection to its substance.

This concept operates along four analytically distinct dimensions. The first is testimonial fabrication, the production of communicative artefacts that perform testimony without originating in any act of witnessing. The second is epistemic extraction, the harvesting of the

visual record of suffering as training data, reproducing colonial logics of resource extraction computationally. The third is delanguaging, the elimination of situated meaning-making from humanitarian representation. The fourth is stereoscopic automation, the algorithmic reproduction at scale of colonial visual tropes that erase pluriversal self-representation.

I do not offer algorithmic witnessing as a comprehensive theory but as a targeted conceptual intervention designed to analyse the specific reconfiguration of communicative and colonial relations that synthetic humanitarianism produces. Its value will be tested in the analysis that follows.

Methodology

This article employs critical conceptual analysis, a non-empirical method in which theoretical propositions are developed, refined, and tested through the systematic examination of illustrative cases (Jaakkola, 2020). Conceptual articles contribute to their field by synthesising relevant theoretical traditions and proposing integrative frameworks that reframe existing phenomena (Jaakkola, 2020, p. 19). The cases in this article function not as a representative sample from which empirical generalisations can be drawn but as an empirical anchor for conceptual elaboration, providing concrete instances against which the analytical adequacy of a proposed framework can be assessed. This approach combines what Jaakkola (2020) identifies as two templates for conceptual research: theory synthesis, in which multiple theoretical lenses are integrated to develop new conceptual tools for an emergent phenomenon, and typology, in which the synthesised framework produces analytically distinct categories that map the variants of that phenomenon (Jaakkola, 2020, p. 22). It is consistent with theoretically driven work published in this journal, including decolonial framework development for assessing AI and creativity (Arora, 2024), conceptual framework building for digital polarization (Esau et al., 2025), and multi-

dimensional framework development for multimodal disinformation (He et al., 2026). The primary contribution is the development of conceptual tools rather than the production of generalisable empirical findings.

The corpus consists of thirteen cases of synthetic humanitarianism: instances in which a humanitarian organisation or campaign actor deliberately deployed computationally generated imagery to perform communicative functions historically reserved for documentary media. Each case represents an occurrence of the phenomenon the article theorises rather than a sample drawn from a pre-existing population. The cases were assembled through iterative monitoring of multiple source types over the period 2022 to 2025, with documentation continuing through March 2026. Sources monitored included humanitarian sector publications (Bond, Third Sector, The New Humanitarian), mainstream media coverage, academic literature (including the Lancet Global Health commentary by Alenichev et al., 2023), organisational websites and press releases, and social media platforms. Targeted searches combined terms such as “AI-generated,” “Midjourney,” and “generative AI” with “charity,” “humanitarian,” “NGO,” and “fundraising.” Initial cases identified through media coverage were supplemented through citation chains in academic and journalistic sources and ongoing monitoring as the practice spread. Over fifty primary and secondary sources were consulted across the thirteen cases.

Cases were included if they met four conditions: the image was generated using a generative AI system; the image was deployed in a humanitarian communication context (fundraising, advocacy, awareness, or campaign material); the use was deliberate rather than accidental or satirical; and the case was publicly documented through media reporting, organisational disclosure, or independent investigation. Cases were excluded if the AI-generated imagery was produced for satire, parody, or protest commentary not claiming humanitarian

authority, or if the use was fraudulent rather than deliberate institutional communication. The Cambodia charity scam cases, excluded on this basis, are discussed in the analysis as a secondary reference point illustrating the downstream consequences of normalised synthetic imagery. Exhibit A-i (Case 3) tests the boundary of these criteria because it was produced by an arts organisation rather than a humanitarian agency, but it was included because it performs humanitarian testimonial functions and was exhibited in humanitarian advocacy contexts (Immigration Museum Melbourne, PhotoVogue Festival). The corpus does not claim comprehensiveness. It captures the range of contexts, actors, and communicative functions through which synthetic humanitarianism has manifested in its early period. Full details of all thirteen cases, including organisation, date, AI tool, communicative context, and primary documentation sources, are provided in Table 2.

The relationship between framework development and case analysis was iterative. The theoretical framework was developed through engagement with witnessing theory, decolonial scholarship, and technocolonialism studies. The four dimensions of algorithmic witnessing emerged through repeated reading of the cases against these theoretical resources. Each case was examined to identify which dimensions were most analytically salient. Cases were then grouped by the dimensions they most productively illustrated, and the analysis was structured accordingly. The analysis does not apply all four dimensions mechanically to every case but identifies the most productive dimensions in each instance and examines how they interact across the corpus. This iterative procedure is consistent with critical conceptual analysis in which theoretical propositions and empirical material develop in dialogue. Figure 1 maps this process.

Figure 1. Analytical process: from theoretical synthesis to framework application

After Jaakkola (2020): theory synthesis + typology

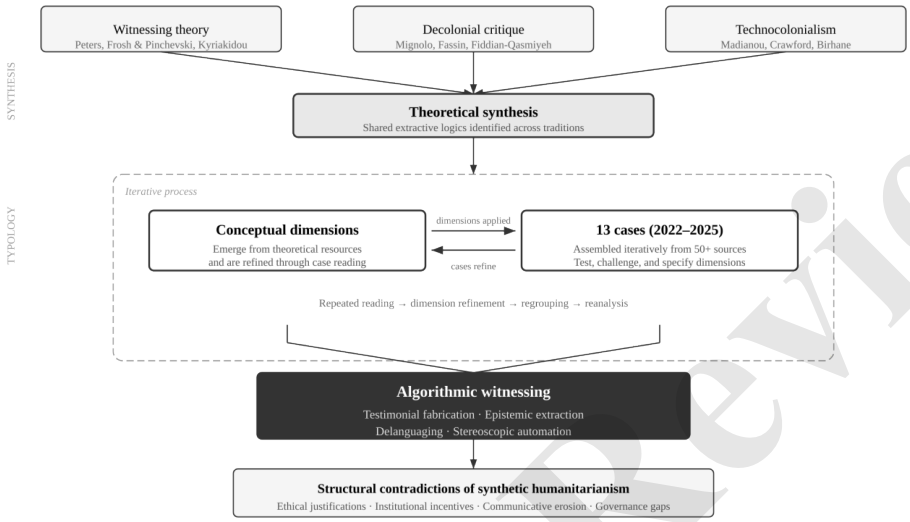

Figure 1. Analytical process: from theoretical synthesis to framework application (after Jaakkola, 2020).

Algorithmic witnessing in practice

Testimonial fabrication

Synthetic humanitarian imagery performs the communicative function of testimony while lacking the epistemic ground on which testimony depends. Across the corpus, every case presents the spectator with an image that issues a moral demand to recognise suffering and respond. Yet in no case does the image originate in an encounter between a recording presence and an actual state of affairs. What varies is the distance between the fabricated image and the reality it claims to represent, and this variation is analytically productive because it reveals that testimonial fabrication operates across a spectrum rather than as a binary condition.

At one end of this spectrum, real events of documentable violence occurred but the image depicting them is computationally generated. Amnesty International Norway's 2023 Colombia campaign (Case 2) depicted abuses during the 2021 national strike using Midjourney-generated imagery justified on protective grounds. The fabrication was exposed through material errors, incorrect flag colours and police uniforms, that documentary photography would not produce. A photograph made in the presence of the event necessarily contains the objects that were there. The synthetic image, generated from generic visual data, lacks this indexical constraint and defaults to approximation. Testimonial fabrication is thus detectable precisely where the synthetic image fails to simulate the constraints that documentary media inherit from the situated encounter.

At the other end of the spectrum, Exhibit A-i (Case 3) presents the most complex configuration. AI images were generated from survivor testimony from Australian offshore detention facilities where documentary photography was structurally impossible due to government policy. Survivors participated in workshops that shaped the generation process. Even here, the dimension applies. The final images are computational translations of verbal accounts, mediated by a system whose visual vocabulary was formed from training data bearing no relation to the conditions described. The testimonial form is preserved. The testimonial act has not occurred. Between these two anchoring cases, the corpus contains entirely fictional subjects presented as testimony to real conditions (Cases 4 and 9) and sanitised depictions of real violence designed for platform circulation (Case 11). In every instance, the witnessing relation (Peters, 2001, p. 709) is maintained in communicative form while its epistemic foundation is removed.

Epistemic extraction

The fabrication of testimony is not a production from nothing. It depends on prior extraction. The dimension of epistemic extraction examines what was taken, from whom, and under what conditions of consent and benefit, in order for synthetic humanitarian imagery to be produced.

The Børns Vilkår campaign (Case 4) illustrates extraction as harm with particular clarity. Generative AI models capable of producing photorealistic images of bruised children must have been trained on datasets containing real images of child abuse, since no other source of such imagery at scale exists (Herrie & Philipsen, 2026). The campaign's ethical justification, that 'no children were harmed' because the depicted children are fictional, collapses under this analysis. The harm was not avoided. It was displaced backward in time, from the moment of image production to the moment of data extraction, when technology companies scraped real images of real abuse into training datasets without consent (Madianou, 2019, p. 8). The synthetic image of a fictional child displayed at a Danish train station is a computational derivative of documented harm to actual children.

This extraction operates not only at the level of individual images but at the level of representational logic itself. Systematic testing has demonstrated that Midjourney reproduces white-saviour and Black-suffering pairings even when explicitly prompted to invert them (Alenichev et al., 2023). The model does not treat the prompt as an instruction to be followed. It reproduces the statistical regularities of its training data, which encode decades of humanitarian visual culture in which whiteness is associated with agency and Blackness with suffering. What has been extracted is not merely a set of images but an entire representational architecture, the colonial visual logic of the humanitarian sector compressed into model parameters and reproduced as default output (Madianou, 2019, p. 2).

The Charity Right A/B test (Case 5) brings the economic dimension of extraction into focus. The dimension of epistemic extraction reveals that by demonstrating near-identical fundraising returns for AI-generated and real images, the test established that from an institutional perspective real persons are economically redundant as subjects of humanitarian representation. The technocolonial cycle of extraction and dependency (Madianou, 2019, p. 2) is functionally closed. The communities whose documented suffering provided computational raw material are no longer needed even as subjects of their own representation. The mass proliferation of AI-generated poverty imagery on stock platforms (Case 12) represents the industrialisation of this logic, making extracted and stereotyped humanitarian imagery available at lower cost and unlimited volume.

Delanguaging

The structural norm across the corpus is the complete absence of affected communities from the image production process. In twelve of thirteen cases, the prompt-to-image pipeline runs from institutional actors to generative AI systems with no point at which the situated knowledge of affected communities enters. The dimension of delanguaging examines the consequences of this absence. When situated voice is structurally excluded from the production process, the witnessing relation loses not only its content but its communicative ground: the capacity of those witnessed to participate in constituting the testimony that claims to speak for them.

The WHO's use of AI-generated avatars to speak on behalf of survivors of wartime sexual violence (Case 10) represents delanguaging in its most extreme form. The avatar does not depict a survivor but fabricates a speaking subject, performing the communicative act of testimony while replacing the survivor's embodied voice, their specific language, their silences and hesitations, with an algorithmic production. What has been called languaging, the situated

and bodily practice of making meaning from one's own geopolitical location (Mignolo & Walsh, 2018, p. 135), is not absent but overwritten by a synthetic proxy that claims the communicative authority belonging to the person it replaces.

The Charity Right A/B test (Case 5), analysed above for its economic implications under epistemic extraction, carries a further consequence under this dimension. The institutional lesson of equivalent fundraising returns is not merely that AI imagery is cheaper but that the situated encounter, the photographer's presence, the subject's participation, the field team's contextual knowledge, adds no measurable communicative value. Delanguaging is thereby validated as economically rational rather than recognised as ethically problematic. This transforms the absence of local voice from an incidental feature of early experimentation into an institutionally authorised practice.

Exhibit A-i (Case 3), analysed above as the most complex instance of testimonial fabrication, is also the case that most directly tests the delanguaging dimension. Survivor participation in workshops represents a genuine attempt to reintroduce situated voice into the production pipeline. Yet even here, the algorithmic layer between verbal testimony and visual output imposes its own representational logic. The AI system's visual vocabulary was formed from training data unrelated to the specific conditions described. Survivor participation mitigates delanguaging but does not overcome it, because the generative system retains control over the visual grammar through which testimony is rendered.

Stereoscopic automation

When delanguaging removes situated voice from the production process, what fills the resulting representational void is not neutral imagery but the automated reproduction of colonial visual

logic. The dimension of stereoscopic automation examines what tropes generative AI systems produce and whether pluriversal modes of representation are accommodated or excluded.

The empirical evidence is direct. Systematic testing has shown that Midjourney defaults to colonial visual pairings regardless of prompt specificity (Alenichev et al., 2023). The automation operates at two analytically distinct levels. The first is the reproduction of established humanitarian stereotypes. AI-generated poverty imagery on stock platforms (Case 12) and in campaigns such as Plan International's anti-child marriage initiative (Case 7) reproduces the iconography of passive suffering, emaciated children, and barren landscapes that critical scholars have identified and challenged for decades (Manzo, 2008, pp. 632–643; Dogra, 2012). The difference is that this iconography is no longer the product of individual representational choices by photographers or campaign designers but of systems that have encoded colonial tropes as statistical defaults and reproduce them at industrial scale.

The second level is the production of new post-humanitarian visual forms (Chouliariaki, 2012) that serve institutional and platform logics rather than representational accuracy. The "All Eyes on Rafah" image (Case 11), analysed above for its testimonial fabrication, reveals under this dimension a specifically post-humanitarian form of stereoscopic automation. The trope it automated was not poverty or victimhood but what I call sanitised solidarity, a visual form optimised for frictionless platform sharing rather than confrontation with suffering. The image contained no bodies, no blood, no specificity. It was designed, whether by intention or by the logic of the generative system, to circulate rather than to disturb. The Amnesty Colombia images (Case 2) reveal stereoscopic automation through failure rather than success, defaulting to generic "Latin American protest" tropes rather than the specific visual conditions of the 2021 national strike. In both cases, the pluriversal reality of particular events in particular places is foreclosed

by a model that substitutes categorical defaults for situated specificity (Mignolo & Walsh, 2018, p. 113).

The four dimensions as system

The analytical separation of these dimensions has been necessary for conceptual clarity, but the cases demonstrate that they operate simultaneously and reinforce one another. Two structural linkages are especially significant. Testimonial fabrication and delanguaging are linked because the elimination of situated representational capacity is what enables fabricated testimony. When no situated encounter informs the image, fabrication meets no resistance from substance. Epistemic extraction and stereoscopic automation are linked because the extraction of colonial visual archives is what produces the statistical regularities that models reproduce as default tropes. The Børns Vilkår campaign (Case 4) illustrates this systemic operation. It simultaneously involves testimonial fabrication (fictional children presented as testimony to real abuse), epistemic extraction (training data likely containing real child abuse images), delanguaging (no affected children or their advocates participated), and stereoscopic automation (the images reproduce normative Western visual tropes of childhood vulnerability). Algorithmic witnessing, as a conceptual framework, is designed to hold these dimensions together analytically rather than treating any one of them as sufficient to account for what synthetic humanitarianism does.

The structural contradictions of synthetic humanitarianism

The research question this article posed asked what conceptual framework is required to analyse the ethical and communicative implications of deploying generative AI imagery in humanitarian communication. The framework of algorithmic witnessing provides the required vocabulary. Applied to thirteen cases, it revealed that synthetic humanitarian imagery fabricates testimony while eliminating its epistemic ground, depends materially on the extraction of visual data

without consent, systematically eliminates local voice and situated knowledge, and automates the reproduction of colonial visual tropes. These four problems form a mutually reinforcing system. Taken together, they demonstrate that synthetic humanitarianism does not merely widen the gap between experience and discourse that Peters (2001, p. 709) identified as constitutive of witnessing but replaces that gap with a computational process that simulates testimonial form without testimonial substance. What follows examines the structural contradictions this system produces.

The most immediate implication concerns the ethical justifications that humanitarian organisations have offered for adopting generative AI imagery. Multiple organisations in the corpus, including Amnesty International (Case 2), Børns Vilkår (Case 4), the United Nations (Case 9), and the WHO (Case 10), framed AI-generated imagery as an ethical choice designed to protect the identities or dignity of vulnerable people. The framework of algorithmic witnessing reveals that these justifications resolve one ethical problem by creating another. The decision not to use real images of identifiable persons addresses legitimate concerns about consent, safety, and representational harm. But the synthetic images that replace them are produced by systems trained on data extracted without the consent of those depicted in that data, and they reproduce the colonial visual logic encoded in training datasets without the knowledge or participation of the communities they claim to represent. The Børns Vilkår case is the sharpest illustration. The claim that “no children were harmed” because the depicted children are fictional collapses when the production process likely depended on training data containing real images of child abuse (Herrie & Philipsen, 2026). The harm was not eliminated. It was displaced from the visible surface of the campaign to the invisible infrastructure of the model.

These ethical justifications operate within the logic of the zero point (Mignolo, 2009, p. 160). The humanitarian institution positions itself as the arbiter of ethical representation, determining what counts as consent, what constitutes harm, and what protective measures are sufficient, while the communities in question remain structurally excluded from this determination. The consent that is obtained, institutional ethics approval, organisational sign-off, legal clearance, is not the consent that matters, that of those whose data was extracted and whose likeness is synthetically reproduced.

The second implication concerns institutional incentive structures. The Charity Right A/B test (Case 5) translated the question of synthetic imagery from an ethical debate into an economic calculation. By demonstrating near-identical fundraising returns for AI-generated and real photographs, the test established that synthetic imagery is not merely defensible but rational. It is cheaper, faster, legally uncomplicated, and equally effective. If these conditions hold, the displacement of documentary photography and local voice will proceed through budget decisions rather than policy decisions. UNICEF Switzerland's decision (Case 6) to reserve AI for "visions" while retaining real photography for frontline documentation represents an institutional attempt to manage this boundary. But if financial returns are equivalent regardless of whether the image is situated or synthetic, the incentive to maintain the boundary weakens with every budget cycle. The normalisation of synthetic humanitarianism should be understood as a structural tendency within the political economy of humanitarian communication rather than as a series of individual organisational choices. The framework specifies why: it is because delanguaging adds no measurable cost and stereoscopic automation produces imagery indistinguishable in fundraising terms from documentary photography that the economic logic favours synthetic production.

The third implication extends beyond humanitarian organisations to the communicative environment in which they operate. The Cambodia charity scam cases, excluded from the primary corpus because they involved fraud rather than deliberate humanitarian communication, demonstrate that the legitimate organisational use of AI-generated humanitarian imagery creates conditions that enable fraudulent imitation. When audiences are habituated to synthetic humanitarian imagery through exposure to campaigns by recognised organisations, the epistemic infrastructure for distinguishing legitimate appeals from fraudulent ones is weakened. The proliferation of synthetic media undermines what Gregory (2021, pp. 1–2) defines as "authenticity infrastructure," the practices and institutions through which audiences assess the credibility of visual testimony. The epistemic challenges extend beyond the humanitarian sector. Synthetic media generates uncertainty that erodes trust not only in specific images but in visual evidence as such (Chesney & Citron, 2019). The visual grammar of humanitarian need becomes available to any actor with access to a generative AI system, and the markers that previously distinguished institutional credibility from opportunistic exploitation become unreliable.

Preliminary research conducted at the University of East Anglia provides empirical support for this concern (University of East Anglia, 2026). When AI-generated images appeared in humanitarian communications, public discussion shifted from the humanitarian cause to debate about AI ethics, with only twenty percent of comments engaging with the actual crisis the campaign addressed. Synthetic humanitarianism creates a self-undermining dynamic. The more humanitarian organisations adopt AI-generated imagery, the more they degrade the communicative environment on which their appeals depend. Trust in humanitarian visual testimony, already weakened by decades of compassion fatigue and post-humanitarian

scepticism (Chouliaraki, 2013, p. 50), is further eroded when the distinction between documentary and synthetic imagery becomes indeterminate.

The fourth implication concerns governance. The dominant regulatory response is disclosure, the requirement that AI-generated images be labelled as such. The framework of algorithmic witnessing reveals that this response is necessary but insufficient. Labelling addresses testimonial fabrication but not the other three dimensions. It does not resolve epistemic extraction, delanguaging, or stereoscopic automation. Disclosure-based governance misrecognises the problem as epistemological, a gap between appearance and reality that transparency can close, when the framework shows it is simultaneously extractive, representational, and colonial. The right to the "in-visible" (Mignolo, 2021, p. 205) and the structural recentring of represented communities as agents of their own representation remain outside the scope of any governance mechanism premised on labelling alone. Exhibit A-i (Case 3) offers a partial model of co-creation, but even here the survivors participated in shaping the images without controlling the system that produced them.

Several limitations of this article should be acknowledged. The corpus captures the early period of synthetic humanitarianism, from 2022 to 2025, and practices are evolving rapidly. New cases, new institutional policies, and new technological capabilities may alter the dynamics analysed here. The framework of algorithmic witnessing was developed for cases of deliberate organisational use and does not directly address grassroots or activist uses of generative AI in humanitarian contexts. The "All Eyes on Rafah" image (Case 11), produced by Malaysian AI artists outside any institutional structure, complicates the technocolonial framing because it originates in the Global South and was created for solidarity rather than fundraising. The framework's four dimensions illuminate aspects of this case, particularly testimonial fabrication

and stereoscopic automation, but the technocolonial analysis of epistemic extraction and delanguaging maps less cleanly onto non-institutional production. Finally, the four dimensions were developed through theoretical analysis and critical conceptual engagement with the case corpus. They have not been tested through audience reception research examining how spectators negotiate the testimonial status of synthetic imagery, nor through ethnographic investigation of the organisational decision-making processes through which synthetic imagery is adopted. These represent necessary directions for future inquiry.

Conclusion

This article has argued that synthetic humanitarianism is not an aberration within humanitarian communication, nor a misuse of an otherwise neutral technology, but an intensification of extractive and colonial logics that have structured humanitarian communication since its institutional emergence. The testimonial model of witnessing conceals a colonial structure in which the humanitarian spectator observes from the zero point while the suffering other is rendered available for moral consumption. Humanitarian visibility operates as an extractive regime that produces appearance and disappearance according to institutional need. Generative AI does not introduce these logics into humanitarian communication. It completes a cycle in which the suffering other is first extracted as training data, then replaced by their own algorithmic reproduction, and finally denied any capacity to contest or reframe their representation. The concept of algorithmic witnessing names this condition and specifies the four analytically distinct dimensions through which it operates.

I argue that the most consequential implication of this analysis is not epistemological but political. The dominant critical response to synthetic humanitarianism has focused on the loss of the indexical referent, on the fact that AI-generated images are not “real” and that their use

undermines the evidentiary basis of humanitarian appeals. This concern is valid but insufficient. What is at stake is not primarily the truthfulness of the image but the elimination of the last structural point at which affected communities could intervene in their own representation. Documentary photography, however asymmetrical and exploitative its history within humanitarian communication, required encounter. A photographer had to be present. In better practice, local image-makers documented their own communities. The encounter was unequal and institutionally mediated, frequently complicit in the very dynamics of extraction and stereotyping that this article has analysed. But it existed, and its existence meant that the represented subject was at minimum a necessary participant in the production of the representation. Synthetic humanitarianism eliminates this necessity. The communities on whose behalf humanitarian organisations claim to act are no longer required at any stage of the communicative process that represents their suffering, completing the extractive architecture that the theoretical framework has described.

Several directions for future research follow from these limitations. Studies examining how spectators negotiate the testimonial status of synthetic humanitarian imagery would complement the theoretical analysis offered here. The “All Eyes on Rafah” case has demonstrated that generative AI is being used by grassroots actors outside Northern institutional structures, and investigation of these uses would test and likely complicate the technocolonial framing this article has proposed. Most urgently, research is needed on how affected communities themselves perceive and respond to synthetic representations produced on their behalf, and on how refugees and displaced communities are already producing their own counter-narratives through digital media (Marino, 2024, pp. 5–8).

The suffering body in humanitarian communication was always mediated; it is now produced. The conceptual tools that communication scholars use to analyse humanitarian practice must be adequate to this shift, or they will fail those whose capacity to represent themselves is at stake.

AI use declaration

All theoretical arguments, analytical claims, and scholarly judgements in this article are entirely my own. I used AI tools to assist with initial coding and cataloguing of the case corpus, structural organisation of early drafts, and proofreading and prose refinement during the writing process. As a disabled and neurodivergent researcher, I employed AI as an assistive technology to address these specific barriers.

Acknowledgment

I am deeply grateful to Lilie, whose inspiration has sustained and shaped my work across two decades.

References

Alenichev, A., Kingori, P., & Peeters Grietens, K. (2023). Reflections before the storm: The AI reproduction of biased imagery in global health visuals. The Lancet Global Health, 11(10), e1496–e1498.

Arora, P. (2024). Creative data justice: A decolonial and indigenous framework to assess creativity and artificial intelligence. Information, Communication & Society, 28(13), 2231–2247. https://doi.org/10.1080/1369118X.2024.2420041

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623.

Birhane, A., & Prabhu, V. U. (2021). Large image datasets: A pyrrhic win for computer vision? In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1537–1547).

Boltanski, L. (1999). Distant suffering: Morality, media and politics. Cambridge University Press.

Butler, J. (2009). Frames of war: When is life grievable? Verso.

Chesney, R., & Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107, 1753–1786. https://doi.org/10.15779/Z38RV0D15J

Chouliaraki, L. (2006). The spectatorship of suffering. Sage.

Chouliaraki, L. (2013). The ironic spectator: Solidarity in the age of post-humanitarianism. Polity.

Couldry, N., & Mejias, U. A. (2019). The costs of connection: How data is colonizing human life and appropriating it for capitalism. Stanford University Press.

Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

Dogra, N. (2012). Representations of global poverty: Aid, development and international NGOs. I.B. Tauris.

Esau, K., Choucair, T., Vilkins, S., Svegaard, S. F. K., Bruns, A., O'Connor-Farfan, K. S., & Lubicz-Zaorski, C. (2025). Destructive polarization in digital communication contexts: A critical review and conceptual framework. Information, Communication & Society, 28(8), 1499–1520. https://doi.org/10.1080/1369118X.2024.2413127

Fassin, D. (2012). Humanitarian reason: A moral history of the present. University of California Press.

Fehrenbach, H., & Rodogno, D. (Eds.). (2015). Humanitarian photography: A history. Cambridge University Press.

Fiddian-Qasmiyeh, E. (2019). The changing faces of UNRWA: From the global to the local. Journal of Humanitarian Affairs, 1(1), 28–41.

Frosh, P., & Pinchevski, A. (Eds.). (2009). Media witnessing: Testimony in the age of mass communication. Palgrave Macmillan. https://doi.org/10.1057/9780230235762

Gregory, S. (2021). Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism. Journalism, 23(3), 708–729. https://doi.org/10.1177/14648849211003313

He, K., Wildfeuer, J., Liu, J., & Guo, J. P. (2026). Diagnosing multimodal disinformation: The integrative 'Three-M' framework of multimodality, manipulation, and malintent.

Information, Communication & Society, 1–23. https://doi.org/10.1080/1369118X.2026.2645883

Jaakkola, E. (2020). Designing conceptual articles: Four approaches. AMS Review, 10(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0

Heller, C. (2014). Post-Fordist capitalism and the media. Pluto Press.

Herrie, M. B., & Philipsen, L. (2026). 'No children were harmed': Aesthetic infrastructures and ethical implications of AI-generated portraits. Visual Studies, 1–12.

Kyriakidou, M. (2015). Media witnessing: Exploring the audience of distant suffering. Media, Culture & Society, 37(2), 215–231.

Madianou, M. (2019). Technocolonialism: Digital innovation and data practices in the humanitarian response to refugee crises. Social Media + Society, 5(3), 1–13.

Manzo, K. (2008). Imaging humanitarianism: NGO identity and the iconography of childhood. Antipode, 40(4), 632–657.

Marino, S. (2024). Refugees' storytelling strategies on digital media platforms: How the Russia–Ukraine war unfolded on TikTok. Social Media + Society, 10(3), 20563051241279248.

Messingschlager, T. V., & Appel, M. (2025). Algorithmic bias in image-generating artificial intelligence: Prevalence and user perceptions. Information, Communication & Society. Advance online publication. https://doi.org/10.1080/1369118X.2025.2584146

Mignolo, W. D. (2009). Epistemic disobedience, independent thought and decolonial freedom. Theory, Culture & Society, 26(7–8), 159–181.

Mignolo, W. D. (2021). The politics of decolonial investigations. Duke University Press.

Mignolo, W. D., & Walsh, C. E. (2018). On decoloniality: Concepts, analytics, praxis. Duke University Press.

Moeller, S. D. (1999). Compassion fatigue: How the media sell disease, famine, war and death. Routledge.

Ong, J. C. (2014). “Witnessing” or “mediating” distant suffering? Ethical questions across moments of text, production, and reception. Television & New Media, 15(3), 179–196. https://doi.org/10.1177/1527476412454687

Ongenaert, D., & Soler, C. (2024). Beyond victim and hero representations? A comparative analysis of UNHCR’s Instagram communication strategies for the Syrian and Ukrainian crises. Journal of Refugee Studies, 37(2), 286–306.

Park, Y. S. (2024). White default: Examining racialized biases behind AI-generated images. Art Education, 77(4), 36–45. https://doi.org/10.1080/00043125.2024.2330340

Peters, J. D. (2001). Witnessing. Media, Culture & Society, 23(6), 707–723. https://doi.org/10.1177/016344301023006002

Richardson, M. (2024). Nonhuman witnessing: War, data, and ecology after the end of the world. Duke University Press.

Robertson, A., & Schaetz, N. (2021). Moving people: Proper distance and global news coverage of migration in 2019. Global Media and Communication, 17(3), 321–343. https://doi.org/10.1177/17427665211039966

Seu, B., & Orgad, S. (2014). The mediation of humanitarianism: Toward a research framework. Communication, Culture & Critique, 7(1), 6–36. https://doi.org/10.1111/cccr.12036

Seuferling, P., & Leurs, K. (2021). Histories of humanitarian technophilia: How imaginaries of media technologies have shaped migration infrastructures. Mobilities, 16(5), 670–687.

Silverstone, R. (2007). Media and morality: On the rise of the mediapolis. Polity.

Sontag, S. (2003). Regarding the pain of others. Farrar, Straus and Giroux.

University of East Anglia. (2026, March). New research warns charities against AI ‘shortcut to empathy.’ https://www.uea.ac.uk/about/news/article/new-research-warns-charities-against-ai-shortcut-to-empathy

Tables

[Table 1. The four dimensions of algorithmic witnessing. See separate file.]

[Table 2. Corpus of synthetic humanitarian communication cases, 2022–2025. See separate file.]

Table 1. *The four dimensions of algorithmic witnessing.*

Dimension	Theoretical origin	Definition	Analytical focus
Testimonial fabrication	Movement 1: Peters (2001), Mignolo (2009)	The production of communicative artefacts that perform the function of testimony without originating in any act of witnessing. The image addresses spectators as witnesses and issues a moral demand, but no event was recorded and no presence registered an actual state of affairs	Whether the communicative artefact maintains any traceable relation to an originating encounter with the real
Epistemic extraction	Movement 2: Fiddian-Qasmiyeh, Madianou (2019), Crawford (2021), Birhane & Prabhu (2021)	The harvesting of the visual record of suffering as training data for generative AI systems, reproducing colonial logics of resource extraction at the computational level. The faces, bodies, and contexts of real persons are converted into computational capital without consent or compensation	Whose data was extracted, under what conditions, and what asymmetries of benefit and control the extraction produces
Delanguaging	Movement 3: Mignolo (2021), Mignolo & Walsh (2018), Malkki (1996)	The elimination of situated, embodied meaning-making from the process of humanitarian representation. The prompt-to-image pipeline severs the connection between the represented and the conditions of their own representability, replacing languaging with command	Whether any situated knowledge, local voice, or embodied encounter informed the production of the representation
Stereoscopic automation	Movement 3: Manzo (2008), Dogra (2012), Mignolo & Walsh (2018)	The algorithmic reproduction at scale of colonial visual tropes that foreclose pluriversal modes of self-representation. Generative AI models trained on historically biased Western datasets encode a Western imagination of crisis as the default grammar of humanitarian imagery	What visual tropes the synthetic image reproduces, and whether pluriversal modes of representation are foreclosed or accommodated

Note. Each cocnept emerges from a specific movement in the theoretical framework and provides an analytical lens for examining the case corpus. The dimensions are analytically distinct but operate simultaneously in any given instance of synthetic humanitarianism.

Table 2. *Corpus of synthetic humanitarian communication cases, 2022–2025.*

#	*Year*	*Organisation*	*Crisis / context*	*Purpose of AI imagery*	*Outcome*
1	2022	Unnamed Toronto charity	Poverty and humanitarian aid, Toronto	Deployed AI-generated images in advertising campaign to raise funds. Among the earliest documented charity uses of generative AI for fundraising	Covered by Toronto Star as emerging trend. Raised questions about transparency and donor trust
2	Apr 2023	Amnesty International (Norway)	Colombia national strike (2021); police brutality and sexual violence against protesters	Generated Midjourney images to illustrate documented abuses while protecting protester identities from state retaliation. Images labelled as AI-generated and used on social media	Immediate backlash from photojournalists and media scholars. Images contained errors (incorrect flag colours, outdated uniforms). Images withdrawn. Became sector reference case
3	Apr 2023	Maurice Blackburn & Howatson+Company	Australian offshore detention; conditions on Nauru and Manus Island where cameras were banned	Created 134 AI images from 300+ hours of survivor testimony, co-produced with detainees in workshops, to provide first visual account of detention conditions	Exhibited at Immigration Museum Melbourne and PhotoVogue Festival Milan 2023. Presented to Australia's Minister for Immigration. Praised for survivor co-creation; sparked ethical debate
4	Apr 2023	Børns Vilkår (Denmark) & Upland Studio	Child abuse and domestic violence against children in Denmark	Created photorealistic AI portraits of five named fictional children with visible bruises for nationwide awareness campaign displayed at all Danish train stations	Peer-reviewed article (Philipsen, 2025) argued training data likely included real child abuse material. Framing of 'no children were harmed' challenged on these grounds
5	Apr 2023	Charity Right (UK)	Hunger crisis; Sudan and Pakistan food aid programmes	Conducted A/B test during Ramadan 2023 comparing AI-generated image against real photograph in Meta fundraising ads. Initially deployed without AI labels	AI images produced near-identical fundraising revenue to real photographs. Detected by Reddit users. Charity published blog defending approach. Divided sector opinion
6	2023	UNICEF Switzerland & Liechtenstein	Children's rights; envisioning better futures for children	Commissioned 12 AI artists to create billboard images depicting hopeful visions for children. Exhibited at Swiss Economic Forum	UNICEF stated AI would be used selectively for 'visions' while retaining real photography for frontline documentation. Regarded as innovative within sector
7	2023–2025	Plan International (UK)	Child, early, and forced marriage globally	Deployed AI-generated images depicting affected communities as part of anti-child marriage awareness campaign, avoiding use of real children's images	Cited in Lancet commentary and Guardian reporting as example of 'poverty porn 2.0'. Identified as part of growing sector trend
8	2023–2025	WWF (UK and Denmark)	Climate crisis and biodiversity loss	WWF UK commissioned AI-generated Romanticism-style paintings imagining climate futures, exhibited at 180 The Strand London. WWF Denmark created AI-generated ads showing consumer products destroying habitats	WWF UK exhibition described as pioneering. WWF Denmark campaign criticised for using energy-intensive AI tools to promote environmental sustainability
9	2023–2025	United Nations	Sexual violence in conflict; documentation and awareness	Generated AI 're-enactment' images of sexual violence situations for advocacy materials, avoiding use of real survivor images	Reported by The Guardian as part of wider trend. Raised concerns about fabricating imagery of sensitive situations. Part of 100+ synthetic charity images catalogued by Alenichev
10	2023–2025	World Health Organization	Health crises; crops-for-profit impact on public health; wartime sexual violence	Deployed AI-generated images in campaign against crops-for-profit and AI-generated avatars to speak on behalf of survivors of wartime sexual violence	Identified in UEA 'Artificial Authenticity' study. WHO defended approach as protecting vulnerability. Criticised for bypassing consent and using AI to 'speak for' affected populations

11	May 2024	Malaysian AI artists (Zila AbKa / Amirul Shah)	Gaza war; Israeli strike on Rafah refugee camp killing 45+ civilians	Created AI-generated image of tent camps spelling 'All Eyes on Rafah' using Microsoft Image Creator as Instagram story template to raise global awareness.	Most viral AI-generated image: 47+ million Instagram shares. Shared by major celebrities. Criticised for sanitising real horrors and enabling 'slacktivism'. Spawned counter-images
12	2023–2025	Various NGOs via Freepik, Shutterstock, and other platforms	Global poverty, hunger, and refugee crises	Purchased mass-produced AI-generated stock images depicting poverty for fundraising materials. Cheaper and faster than commissioning real photography.	Lancet commentary coined 'poverty porn 2.0'. Researcher Alenichev catalogued 100+ images. Criticised for reinforcing racial stereotypes and colonial tropes. FairPicture published white paper on risks
13	Jan 2024	Unnamed large Western European NGO	Humanitarian crisis in West Asia	Prepared campaign using AI-generated images of West Asian families. Withdrawn before publication over concerns about cultural insensitivity.	Reported in The New Humanitarian (October 2025). Only documented case of institutional self-correction before publication. All other known withdrawals followed public backlash

Note. Cases were included if the image was generated using a generative AI system, was used in a humanitarian communication context (fundraising, advocacy, awareness, or campaign material), the use was deliberate rather than accidental or satirical, and the case was publicly documented through media reporting, organisational disclosure, or independent investigation. Cases are ordered chronologically. Three additional items documented during corpus construction (the UEA 'Artificial Authenticity' sector-wide study by Alenichev et al., the Alenichev et al. Midjourney colonialist bias study published in Lancet Global Health, and the Cambodia AI charity scam cases) were excluded from the primary corpus and are discussed as secondary sources in the analysis and discussion sections respectively.