The Efficiency–Equity Tradeoff in AI-Enabled Humanitarian Information Systems: Toward an Intersectional Governance Audit Framework

Guljannat Huseynli
Old Dominion University
ghuse001@odu.edu

Erika Frydenlund
Old Dominion University
efrydenl@odu.edu

ABSTRACT

Artificial intelligence (AI) systems are increasingly embedded in humanitarian and disaster response, including flood forecasting, satellite-based damage assessment, displacement prediction, and biometric aid registration. While governance debates emphasize fairness and accountability, most systems are optimized for predictive accuracy, speed, and operational efficiency. We map current humanitarian AI deployments and identify a recurring structural pattern: efficiency-driven design can sideline gendered and intersectional vulnerabilities. Through case analyses of flood forecasting, rapid damage assessment, and biometric registration, we examine how optimization goals, units of analysis, and performance metrics determine what counts as relevant harm, and how the abstraction required for AI modeling compresses complex social contexts in ways that limit both contestability and oversight. We introduce a draft Intersectional Governance Audit Framework organized around optimization logic, administrative embedding, and governance alignment to guide evaluation of humanitarian AI design choices upstream of deployment.

Keywords

Humanitarian AI, Humanitarian Management, AI Governance, Feminist AI, Intersectionality

INTRODUCTION: NAMING THE EFFICIENCY–EQUITY TRADEOFF IN HUMANITARIAN AI

Artificial intelligence is now embedded across the humanitarian crisis cycle. Predictive models forecast floods and displacement; satellite imagery estimates damage; biometric systems manage aid eligibility; and chatbots provide health and legal information. Humanitarian actors increasingly invoke "responsible AI," emphasizing fairness, transparency, and oversight. Yet most systems are focused primarily on predictive accuracy, speed, coverage, and logistical efficiency. Equity commitments often remain policy language rather than formal constraints within model design.

This paper identifies a structural tension at the core of humanitarian AI: what we're calling the efficiency–equity tradeoff. ML/AI systems require defined units of analysis, measurable variables, and optimization objectives. Large language models (LLMs), for example, are trained on datasets whose weightings remain black boxes to end-users. In emergency contexts, optimizing for speed and scale can appear morally necessary, but optimization depends on abstraction, and increasingly relies on unknowable training data. Decisions about what to measure, how to abstract, and which harms count determine whose vulnerability the system can see. Because these decisions are embedded in upstream design (in the choice of analytical unit, in feature selection, in the optimization objective), they are rarely legible in system outputs, difficult to contest after deployment, and require instruments that reach into the design pipeline itself.

The problem is not only biased inputs and outputs. Intersectionality – the way overlapping structures of power co-

produce experiences of marginalization that no single axis can capture (Crenshaw, 1991) – is what efficiency-driven humanitarian AI fails to encode within its core abstractions. Feminist scholarship on AI has shown that when such experiences are not built into system design, they are rarely recoverable after the fact (Paton-Romero et al., 2022; Su, 2025). In humanitarian AI, efficiency-driven objectives frequently fail to encode this kind of intersectional vulnerability within their core abstractions. After the 2023 Türkiye–Syria and Afghanistan earthquakes, where most casualties were women and children trapped inside collapsed homes, AI-assisted damage assessment relied on building-footprint classification as its analytical unit (Ramiah, 2024; Yu et al., 2024). Because gender, age, and household composition were not part of the model's abstraction, the disproportionate mortality of women and girls could not be registered when recovery decisions were being made (Beduschi, 2022; O'Connor & Liu, 2023; Su, 2025). A parallel pattern appears in biometric registration systems that enroll at the household level, where intra-household power asymmetries are smoothed out by the unit of analysis itself (Van den Homberg et al., 2020). In both cases, the problem is not a biased input but a structural one, making intersectional vulnerability invisible. Governance debates tend to focus on fairness after deployment, while upstream design choices remain largely unexamined. To build a starting place to explore these ideas, we propose draft *Intersectional Governance Audit Framework* that seeks to highlight abstraction-level design, administrative embedding, and governance alignment in humanitarian AI systems.

Three research questions guide this work: RQ1: What patterns recur across current humanitarian AI deployments in how optimization objectives and units of analysis are defined? RQ2: Through what upstream design choices do efficiency-driven systems render intersectional vulnerability invisible to operational decision-making? RQ3: What governance dimensions would enable scrutiny of these upstream choices, beyond the post-deployment focus of prevailing fairness debates?

Our contribution is threefold: (1) a landscape mapping of humanitarian AI deployments across the crisis cycle (Table 1) that surfaces a recurring efficiency–equity pattern; (2) three illustrative case analyses — Flood Hub, RAPIDA, and BIMS/PRIMES — that trace how upstream design choices render intersectional vulnerability invisible to operational decision-making; and (3) a draft Intersectional Governance Audit Framework intended to guide scrutiny of humanitarian AI design.

AI ACROSS THE HUMANITARIAN CRISIS CYCLE

The integration of artificial intelligence into humanitarian and disaster response has accelerated sharply in recent years (Beduschi, 2022; Haykal, 2025; Lythreatis et al., 2026). Lythreatis et al. (2026) document a significant increase in research activity since 2020, while Haykal (2025) surveying literature from 2001 to 2025, confirms that most activity has concentrated after 2015. The two reviews map applications across overlapping domains: from disaster prediction and displacement forecasting to disease surveillance, psychosocial support, robotics-assisted care, and supply chain logistics. A 2025 EU Emergency Response Coordination Centre review found that AI performs most consistently in standardized, data-heavy crises (floods, wildfires, droughts) but struggles in novel scenarios, low-data environments, and contexts requiring moral judgment (SAPEA, 2025). These are the conditions most common to humanitarian work.

The humanitarian crisis cycle refers to the sequence of humanitarian activities surrounding a crisis event including anticipation, preparedness, response, and recovery (Lythreatis et al., 2026; Sperling & Schryen, 2022). It overlaps substantially with the standard disaster management cycle, with additional emphasis on protection, assistance, and displacement(Beduschi, 2022; Mshelia et al., 2025). Across this cycle, AI deployments fall into two temporal categories and one cross-phase category, reflected in Table 1 as "Pre-Crisis," "Post-Crisis," and "Cross-Cutting." The last denotes systems that operate across all phases rather than a distinct phase in its own right.

Pre-crisis systems include flood forecasting platforms, anticipatory action triggers, predictive displacement models, and hunger mapping tools. Google's Flood Hub, launched in 2022, combines hydrologic simulation with ML to generate riverine flood forecasts up to seven days in advance, extending forecasting capacity into historically data-scarce regions (Google Research, 2026). Similar predictive systems integrate climate, economic, and market indicators to anticipate displacement or food insecurity trends (Lythreatis et al., 2026). Early warning systems have also expanded institutionally. The UN has launched initiatives aimed at integrating AI into the early warning value chain, including the Global Initiative on Resilience to Natural Hazards through AI Solutions and the World Meteorological Organization's EW4All Executive Action Plan (2023–2027) (Salio et al., 2025; WMO, AICCRA Concept Notes/2023).

Post-crisis applications focus on damage assessment, logistics coordination, and aid allocation. Satellite imagery and computer vision models estimate infrastructure destruction when field access is constrained. The UNDP Crisis Bureau's RAPIDA tool, for example, integrates satellite imagery, social media signals, and night-light data for rapid impact assessment, and was deployed following the 2023 Herat earthquake to estimate damage in areas inaccessible to field teams (Ramiah, 2024). Optimization algorithms support routing and distribution decisions,

while ML tools assist with triage and resource prioritization in health and emergency contexts (Haykal, 2025).

Cross-cutting systems include biometric registration platforms, digital identity infrastructures, predictive dashboards, and chatbot interfaces for information dissemination. The UNDP Crisis Bureau has deployed EVA.ai for expert roster matching across more than 150 offices and a Crisis Risk Dashboard combining near-real-time monitoring with predictive analytics, adapted in contexts such as Sri Lanka and Ecuador (Ramiah, 2024; United Nations, 2024). Biometric systems operated by agencies such as UNHCR and WFP link identity verification to aid distribution across multiple countries (Beduschi, 2022). These infrastructures often function as backbone systems, integrating eligibility, monitoring, and coordination functions across agencies.

Practitioners caution that these systems do not operate on neutral terrain (Mshelia et al., 2025; Parkes et al., 2022; Pizzi et al., 2020; Su, 2025; Van den Homberg et al., 2020). In many crisis-affected regions, particularly across parts of the Global South, digital infrastructure is uneven, technical capacity is limited, and reliable data are difficult to obtain. As a result, the communities with the greatest need for effective response are often those for whom AI system performance is least certain (Mshelia et al., 2025; Parkes et al., 2022). Beduschi (2022) describes this pattern as "techno-solutionism:" treating deeply rooted social and political vulnerabilities as if they can be resolved primarily through improved technical modeling.

Phase	Tool/System	Use Case	AI Technique	Key Limitations
Pre-Crisis	Google Flood Hub	7-day river flood forecasting for anticipatory action	Hydrologic + inundation ML models	Forecast reliability depends on gauge data; "virtual gauges" in data-scarce regions are less-validated (Nearing et al., 2024)
	WFP HungerMap LIVE	Nowcast food insecurity trends for early humanitarian action	ML predictive analytics (LSTM, XGBoost)	Monitoring gaps affect regional representativeness (Lythreatis et al., 2026)
	OCHA Cyclone Impact Model	Predict cyclone housing damage pre-landfall for anticipatory financing	ML impact model for forecasts + static damage features (often XGBoost)	Accuracy constrained by sub-national historical data and forecast archives (Lythreatis et al., 2026)
	UNHCR Project Jetson	Forecast displacement movements for contingency planning	Supervised ML time-series forecasting	Somalia displacement forecasts built on climate, market, and remittance data; structural violence and gender-specific drivers of flight absent from inputs (Beduschi, 2022)
	IFRC Forecast-based Financing	Release funds for early action based on forecast triggers	ML + meteorological/ market data triggers	Misclassification delays resource release; event-specific accuracy varies (Beduschi, 2022)
	UNDP Crisis Risk Dashboard	Monitor multidimensional risks and forecast crisis hotspots	AI enabled predictive risk analytics + NLP monitoring	Country-specific customization required; automated response triggers not yet operational (Ramiah, 2024)
	Shelter site selection models	Urban emergency shelter site selection using spatial-topology learning	ML model (VAG + RF)	Optimizes urban demand factors; gendered protection needs and intra-household dynamics absent from inputs (Wang et al., 2024)
	Anticipatory action triggers (Philippines)	Trigger cyclone early action using probabilistic housing-damage forecast	ML damage prediction	"Many hands" accountability problem: responsibility dispersed across developers, agencies, and governments (UNOCHA, 2024)
Post-Crisis	UNDP RAPIDA	Rapidly estimate disaster impacts to guide field assessments	GIS + satellite imagery + social media fusion	Validated in Afghanistan; generalizability to conflict zones with restricted satellite access untested (Ramiah, 2024)
	Satellite damage assessment	Map building-damage categories from satellite imagery	Deep learning + computer vision + remote sensing	Performs significantly better on Global North; region-specific retraining essential but rarely resourced (UNOSAT, 2026; Van den Homberg et al., 2020)
	AI-assisted triage and hospital resource allocation	Predict ED acuity for triage support	ML-based triage support	Tested in stable health systems; performance in conflict-affected healthcare unknown (Haykal, 2025)
	Relief distribution optimization	Optimize last-mile routing and supply allocation	Agent-based simulation + metaheuristics	Optimizes cost and speed; equity constraints and gendered access barriers rarely encoded (Lythreatis et al., 2026)
	Social media disaster classification (IFRC)	Classify crisis posts for situational awareness	NLP + computer vision	Digital data overrepresents connected populations; rural, elderly, and displaced communities underrepresented (Beduschi, 2022)
	Humanitarian OpenStreetMap /AIDR	Crowdsourced crisis mapping and message classification	ML + satellite imagery + volunteer classification	Misinformation risk; inconsistent verification; governance for crowdsourced data largely absent (Ziemke et al., 2021)

	mHealth (IRC /Lebanon)	Mobile records and decision support for NCD care	Mobile patient-controlled health records	Provider uptake low (21.7% of consultations); duplication with paper systems undermines adoption (Doocy et al., 2016)
	Telemedicine in conflict zones (WHO EMRO)	Remote consultations when conflict disrupts local care	Store-and-forward + real-time consultation platforms	Bandwidth, equipment, and funding constraints; outcome evaluation largely absent (Parkes et al., 2022)
	CIMA	Vaccine reminders and refugee child immunization tracking	Mobile reminder + tracking	19% reduction in vaccine loss-to-follow-up; scalability beyond pilot undemonstrated (Haykal, 2025)
	HERA Digital Health	WhatsApp-based navigator for prenatal and child vaccination services	Agentic AI chatbot + GIS service locator	Peer-reviewed outcomes limited; scalability beyond Turkey/Syria context undemonstrated; sustainability donor-dependent (HERA Digital, 2026; Sürmeli, 2024)
	UNHCR BIMS / PRIMES	Register refugees and verify identity for protection and assistance	Biometric recognition + digital identity management	Household-head enrollment logic excluding stateless, undocumented, and informally displaced persons (UNHCR, 2026)
	WFP SCOPE (biometric aid verification)	Register households and verify beneficiaries at distribution points	Biometric verification at distribution points	Intra-household power asymmetries invisible; fingerprint-match errors higher for manual laborers and refugees (CALPNetwork, 2022; WFP, 2026)
Cross-Cutting	NLP chatbots	Crisis information and GBV support	LLM-powered conversational AI	Fail to recognize GBV survivors' coded language; data breach risk; no redress mechanisms (Spencer & Masboungi, 2025)
	Wysa, Woebot (mental health chatbots)	Self-guided CBT exercises and check-ins for stress and anxiety.	CBT and mindfulness NLP	Limited multi-language capability; weak data privacy governance; no human oversight for acute crises (Haykal, 2025)
	ZzappMalaria / IBM Watson Health	AI decision-support for malaria larval source management operations	Data-driven planning and mapping for vector-control fieldwork	Equitable deployment in low-infrastructure settings unresolved; training data representativeness unverified (Haykal, 2025)
	Blockchain identity and medical records	Share records across agencies via a tamper-resistant ledger	Blockchain-based identity/records	Field-effectiveness evidence scarce; most deployments are pilots (Hunt et al., 2022)
	Participatory AI / collective crisis intelligence	Community validation of data and model training	Community-curated ML + human-in-the-loop	Promising for bias mitigation; rarely scaled beyond pilot; localization under-implemented (Kathy Peach, 2022)

Table 1. AI Applications Across the Humanitarian Crisis Cycle

Table 1 maps a selection of AI tools and systems across the crisis cycle to document representative deployments. It was compiled through a targeted review of three source types: recent systematic reviews of humanitarian AI, peer-reviewed empirical studies of specific deployments (Haykal, 2025; Lythreatis et al., 2026; Ortega-Bolaños et al., 2024; SAPEA, 2025), and operational documentation from implementing organizations (UNHCR, WFP, UNDP, OCHA, IFRC). Tools were included where documentation confirmed field deployment and where coverage across the three crisis-cycle phases could be maintained.

Three cases are carried forward: Google Flood Hub, UNDP RAPIDA, and UNHCR BIMS/PRIMES. They span the three phases, three architectures (predictive modeling, geospatial fusion, biometric verification), and three governance configurations (private–humanitarian partnership, UN agency deployment, UN-operated identity infrastructure), allowing the framework to be tested against meaningfully different systems.

Several patterns are visible across application categories, each reflecting the efficiency–equity tradeoff in a different register. First, systems are most mature in standardized, data-rich domains such as flood forecasting and satellite-based damage assessment (SAPEA, 2025). Second, optimization objectives consistently prioritize predictive accuracy, coverage, and logistical efficiency. Third, documented limitations frequently concern data representativeness, infrastructure gaps, and contextual validation (Beduschi, 2022; Parkes et al., 2022). Less frequently examined in deployment documentation are questions about how systems account for gendered access to technology, intra-household dynamics, trauma-informed modeling, or intersecting forms of vulnerability. Where such considerations appear, they are typically addressed through disaggregation at the reporting stage rather than formalized within model design. Reading through an equity lens, many of the limitations in Table 1 are downstream consequences of efficiency-driven design choices: who is counted, whose context is measurable, whose vulnerability is modeled.

The SAPEA (2025) review calls for systematic risk assessment, data harmonization, and strengthened human oversight, while noting that sector-wide benchmarks and enforcement mechanisms remain underdeveloped. Implementation studies reinforce this gap between technical capability and contextual fit. Doocy et al. (2016)

found an *mHealth* intervention among Syrian refugees in Lebanon, adopted in fewer than a quarter of consultations despite measurable clinical improvements. Parkes et al. (2022) similarly identify limited outcome evaluation and heavy reliance on volunteer providers in conflict-zone telemedicine. The landscape reflects rapid technological integration alongside uneven governance adaptation.

ESTABLISHED GOVERNANCE CRITIQUES AND THEIR LIMITS

In this paper, the efficiency–equity tradeoff names a design-level problem: what AI systems do to intersectional vulnerability through their analytical unit, optimization objective, and constraint settings. Governance refers to the institutional mechanisms meant to hold those design choices accountable. Our argument is that current governance mechanisms engage the downstream symptoms of the structural tension rather than the upstream design choices that produce it.

Governance scholarship on humanitarian AI has consolidated around a recognizable set of concerns. First, researchers documented systematic asymmetries in whose data informs field systems. Beduschi (2022) argues that models trained in data-rich environments are routinely transferred into contexts they were never designed for, reproducing global power. Empirical evidence supports this claim. Weber de Melo et al. (2025) find disaster image classification systems perform significantly better on European imagery than on data from Africa or Asia. Van den Homberg et al. (2020) document accuracy discrepancies in building map datasets used in Tanzania and the Philippines large enough to affect damage assessments. Joshi et al. (2020) show that languages most spoken by the world's population, like African languages, receive only small fraction of NLP research attention.

At the institutional level, the UN High-level Advisory Body on AI (2024) identifies a global governance deficit, noting that the majority of states are absent from major international AI governance initiatives (United Nations, 2024).

A second cluster of critiques centers on opacity and responsibility. Pizzi et al. (2020) argue that humanitarian AI deployments depend heavily on corporate self-regulation, with limited democratic oversight and no clear mechanism for remedy when harm occurs. Lythreatis et al. (2026), reviewing sixty peer-reviewed studies, found limited reporting of explainable AI methods. Van den Homberg et al. (2020) describe a "many-hands problem," in which responsibility for AI-generated decisions disperses across developers, donors, implementing agencies, and government actors. Dirven (2024) argues that fairness-oriented ML efforts often fall into "structural traps." One is the framing trap, which treats technical systems as separate from the social context. Another is the solutionism trap, which recasts complex social problems as technical adjustments. Teodoro et al. (2022) document repeated failures across ethical requirements in deployed systems, reinforcing the concern that governance frameworks often operate at the level of principle without altering underlying deployment logics.

A third line of critique questions ethics frameworks themselves. Comes (2024), building on Mittelstadt (2019)'s, argues that widely cited AI ethics principles in language but diverge sharply in practice. Jobin et al. (2019), mapping 84 global AI ethics guidelines, find broad agreement on values but variation in enforcement and limited Global South participation.

Bietti (2020) describes "ethics washing" as the strategic use of ethical language in the absence of binding oversight, while van Maanen (2022) argues that principle-driven governance distracts from questions of power and political economy as summarized by Lidén (2026). Greene et al. (2021) show how ethics initiatives can incorporate critical language while keeping decision-making authority concentrated among technical and institutional elites.

This literature identifies uneven representation, limited transparency, and weak accountability in humanitarian AI governance. Yet these concerns are typically framed as correctable defects after deployment. Less attention is given on how problems are defined, what is optimized, and which harms fall within model scope. In humanitarian contexts where speed and scale are treated as urgent imperatives, these framing choices shape whose vulnerability becomes actionable.

AN INTERSECTIONAL GOVERNANCE LENS

Prevailing governance critiques of humanitarian AI focus on bias, transparency, and accountability after systems are deployed. Our take on an intersectional feminist lens shifts attention upstream to how problems are defined, categories constructed, units of analysis selected, and outcomes optimized. The issue is not only whether a system performs fairly, but how design choices determine what can be seen, counted, and acted upon in the first place.

Recent gender and AI scholarship clarifies this distinction. Paton-Romero et al. (2022) differentiate between using AI to advance gender equality ("gender by AI") and mitigating gendered harms within AI systems ("gender in AI"). As Scheuerman et al. (2021) argue through the concept of "auto-essentialization," encoding gender as fixed

and binary does more than risk misclassification; it formalizes contested social categories in technical, unseeable form. When such categories become operational, particularly in systems governing access to aid, they shape how individuals are recorded, sorted, and able to contest decisions.

Intersectional feminist analysis complicates representation further. Crenshaw's (1991) concept of intersectionality is that overlapping structures of power — gender, race, class, citizenship, disability, displacement status — co-produce forms of marginalization that no single axis can capture. For humanitarian AI, the methodological implication is that a system registering gender alone, or displacement status alone, cannot represent the vulnerabilities that emerge specifically at these intersections. Aggregated data structures commonly used as the basis of these systems — households, risk scores, population averages — smooth over such intersections, improving efficiency while obscuring inequity. This dynamic reflects the logic of optimization itself, and it is not new to humanitarian data; ML compounds it by making the link between design intent and outcome more opaque and harder to trace. As Su (2025) argues, building on the concept of 'slow violence,' harms accumulate gradually through routine administrative processes; these choices shape eligibility, risk classification, and resource prioritization over time.

We develop a draft Intersectional Governance Audit Framework that recenters who defines system objectives, what is optimized, and which realities remain outside complex, opaque AI design pipelines. Rather than treating speed and scale as neutral imperatives, we treat them as political choices requiring scrutiny. The following section applies this lens to the mapped humanitarian AI landscape.

APPLYING AN INTERSECTIONAL GOVERNANCE LENS TO HUMANITARIAN AI GOVERNANCE: AN INTERSECTIONAL GOVERNANCE AUDIT FRAMEWORK

We now apply the intersectional feminist governance lens to some humanitarian AI systems mapped in Table 1. Rather than evaluating tools solely on performance, we examine how design choices shape what vulnerabilities are incorporated into operational decision-making. Our analysis focuses on three dimensions: (1) optimization objectives, (2) administrative embedding, and (3) governance alignment. Together, these elements direct attention to abstraction-level design decisions—what is measured, how success is defined, and how system outputs structure access and authority within humanitarian workflows. The following cases illustrate how these dimensions operate across different humanitarian AI systems.

Framework Element 1: Optimization & Performance Metrics

This element examines how performance objectives are defined within humanitarian AI systems. Optimization is not neutral; it encodes institutional priorities about what counts as "important" or "success." In humanitarian settings, systems are frequently designed to maximize speed, coverage, predictive accuracy, or logistical efficiency. Our framework evaluates how these objectives are selected, what trade-offs they formalize, and whose experiences are incorporated (or excluded) within performance metrics.

Case 1.

Google Flood Hub optimizes for hydrological accuracy. Its Long Short-Term Memory (LSTM) based model forecasts river discharge and generates 15-meter-resolution flooding maps, evaluated against gauge performance and systems like GloFAS across ~150 countries (Google Research, 2026; Nearing et al., 2024).

Human vulnerability is not part of the model. Gender, disability, displacement, and informal settlement status are absent from training data, and the system excludes urban inundation. What surfaces as "bias" is a consequence of how the system defines its analytical unit, not a correctable input error. The equity gap is structural, not accidental, because success is defined in environmental rather than social terms.

Case 2.

RAPIDA prioritizes speed. It fuses satellite imagery, nighttime lights, and geospatial data to generate rapid building damage estimates when field access is limited, as seen after the Herat earthquake and Storm Daniel (Ramiah, 2024).

No public accuracy metrics compare outputs to ground truth. The unit of analysis is the building footprint, and gender, age, and household dynamics enter only in later human-led stages. In Herat, where most casualties were women and children inside collapsed homes, building-level classification could not register differentiated mortality. Open building datasets that undercount informal settlements further risk minimizing the most vulnerable at the moment recovery funds are allocated. Here too, what surfaces as "bias" is the downstream

signature of upstream choices about what the model measures.

Case 3.

BIMS/PRIMES optimizes for identity uniqueness. Refugees aged five and above are enrolled through fingerprints, iris scans, and facial images linked to a centralized database, with success measured by enrollment coverage across 93 countries (UNHCR, 2023, 2025, 2026).

Public reporting does not include disaggregated failure rates, even though age-related skin changes and fingerprints worn through repeated manual work can reduce ridge clarity, increasing the risk of poor capture, failure to enroll, or false non-matches. Although the system records vulnerability markers, its logic centers on identity certainty. Aid is often distributed at the household level, obscuring intra-household power dynamics. Among Rohingya refugees in Bangladesh, registration data collected for identity cards and access to essential services were later used in repatriation eligibility assessments shared with Myanmar, illustrating how verification infrastructure can shape outcomes beyond aid delivery (HRW, 2021). The system secures identity but does not assess the broader political consequences of that design.

Framework Element 2: Administrative Embedding And Contestability

Humanitarian AI systems rarely function as standalone analytic tools. They are embedded within administrative processes such as eligibility determination, beneficiary targeting, risk classification, and resource allocation. This element evaluates how AI outputs structure access and whether mechanisms exist for contestation and correction.

Case 1.

Flood Hub forecasts feed into automated anticipatory action chains. In Nigeria, GiveDirectly's system uses API-triggered thresholds to release cash transfers within 48-72 hours (Barreras & Kenkuyu, 2025; Google Research, 2026). Although formally advisory, trigger generation requires no human intervention.

Communities that are not included in the model or do not trigger action may have no way to ask why or request a review since feedback channels are designed for API users rather than affected populations. Over time, communities excluded from triggers produce no outcome data and are absent from scaling evidence; absence becomes self-reinforcing.

Case 2.

RAPIDA narrows which areas receive in-person assessment teams, shaping the scope of the Post-Disaster Needs Assessment that determines recovery funding (Ramiah, 2024). Undercounts in building data carry forward into funding decisions without a clear correction stage. Technical auditability does not translate into practical access for affected communities. Stateless and nomadic populations are especially vulnerable, underrepresented in administrative records, undercounted in datasets, and weakly represented in government-led PDNA processes.

Case 3.

Unlike the previous systems, PRIMES gates access directly. A biometric match permits aid; a non-match diverts the individual to secondary review. UNHCR policy states that biometrics alone should not determine exclusion, yet documented cases show refugees losing access during enrollment disputes.

Formal grievance mechanisms exist but are not biometric-specific. Audits by the Office of Internal Oversight Services (OIOS) found that some registration-system synchronization incidents took up to 621 days to resolve (UN, 2025; UNHCR, 2026). People whose fingerprints or other biometric data are difficult for the AI and bio-technology systems to capture may face greater delays in verifying their identity. Repeated verification problems can also reduce their individual visibility in household-based records, making it harder to access protection referrals or resettlement pathways. The cumulative effect is a quiet erosion of individual agency.

Framework Element 3: Governance Alignment and Abstraction Oversight

The third element examines whether existing governance mechanisms address abstraction-level decisions. Current frameworks emphasize bias mitigation, transparency, and compliance. This element evaluates whether governance processes interrogate how units of analysis, objective functions, and constraint settings are defined. It also considers whether oversight mechanisms account for recursive effects as systems shape future data and model updates.

Case 1.

Flood Hub operates under Google's internal AI Principles. Humanitarian partners apply governance frameworks to forecast use, not to forecast design. No external body evaluates Flood Hub's optimization logic, unit of analysis, or exclusion of urban inundation. Responsibility for abstraction choices remains internal to Google; operational consequences are borne by humanitarian actors.

Case 2.

RAPIDA falls under UNDP Social and Environmental Standards and UN data responsibility guidelines. OIOS audits assess compliance, not whether building footprints represent human need or whether source datasets undercount informal settlements. Communities do not participate in design or validation; Post-Disaster Needs Assessments (PDNA) are government-led, leaving affected populations limited opportunities to challenge estimates. Open code does not itself ensure accountability.

Case 3.

PRIMES operates within a comparatively developed governance architecture: Data Protection Policy, Age, Gender and Diversity approach, OIOS audits. These policies and audits evaluate whether the system functions as designed; it does not question whether identity uniqueness should be the primary optimization objective. Complaint mechanisms address individual cases but not cumulative patterns. Structural tensions between individual biometric verification and household-level aid distribution remain outside formal governance review.

CONCLUSION: FROM BIAS TO STRUCTURE IN HUMANITARIAN AI GOVERNANCE

This work in progress argues that prevailing debates on humanitarian AI governance remain too focused on bias, transparency, and post hoc accountability. These concerns matter but do not interrogate how upstream design decisions structure inclusion and exclusion. Across the crisis cycle, systems are optimized for speed, scale, and predictive performance, formalizing units of analysis and objective functions that shape which harms become actionable and which forms of intersectional vulnerability remain peripheral.

Whether in flood forecasting, rapid damage assessment, or biometric registration, abstraction-level choices about what to measure and optimize influence how vulnerability is rendered visible, how access is structured, and how contestation is constrained. These effects emerge through routine administrative embedding and accumulate as systems become institutionalized. We propose a draft Intersectional Governance Audit Framework organized around optimization logic, administrative embedding, and governance alignment. Further research is needed to operationalize these dimensions empirically and to examine how organizational incentives and public–private partnerships shape design priorities. Redirecting attention from bias alone to the politics of abstraction is a necessary step toward more accountable humanitarian AI governance.

AI WORKFLOW DISCLOSURE

AI-assisted tools were used for literature discovery (Elicit, Research Rabbit, Web of Science), thematic synthesis (NotebookLM), and drafting support (ChatGPT 5.2). All analysis, interpretation, and substantive arguments are the responsibility of the authors.

ACKNOWLEDGMENTS

We thank the reviewers, program and local committee members of the ISCRAM conference for their valuable feedback.

REFERENCES

Barreras, F., & Kenkuyu, S. (2025). How AI helped send cash just before floods hit in Nigeria. https://www.givedirectly.org/flood-forecast-ai

Beduschi, A. (2022). Harnessing the potential of artificial intelligence for humanitarian action: Opportunities and risks. *International Review of the Red Cross*, *104*(919), 1149–1169. https://doi.org/https://doi.org/10.1017/s1816383122000261

Bietti, E. (2020). From ethics washing to ethics bashing: a view on tech ethics from within moral philosophy. Proceedings of the 2020 conference on fairness, accountability, and transparency,

Comes, T. (2024). AI for crisis decisions. *Ethics and Information Technology*, *26*(1), 12. https://doi.org/10.1007/s10676-024-09750-0

Crenshaw, K. (1991). Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review*, *43*(6), 1241–1299. https://doi.org/10.2307/1229039

Dirven, E. (2024). The Application of Large Language Models by the Red Cross Red Crescent Movement: Upholding Impartiality and Neutrality.

Doocy, S., Lyles, E., Akhu-Zaheya, L., Oweis, A., Al Ward, N., & Burton, A. (2016). Health Service Utilization among Syrian Refugees with Chronic Health Conditions in Jordan. *PLOS ONE*, *11*(4), e0150088. https://doi.org/10.1371/journal.pone.0150088

Google Research. (2026). *Using AI to make critical flood forecasting information universally accessible*. Google. https://sites.research.google/gr/floodforecasting/

Greene, M. C., Haddad, S., Busse, A., Ezard, N., Ventevogel, P., Demis, L., Inoue, S., Gumm, J.-C., Campello, G., & Tol, W. A. (2021). Priorities for addressing substance use disorder in humanitarian settings. *Conflict and health*, *15*(1), 71.

Haykal, D., Goldust, M., Cartier, H., & Treacy, P. (2025). AI in humanitarian healthcare: a game changer for crisis response [Mini Review]. *Frontiers in Artificial Intelligence*, *8*, 1627773. https://doi.org/1627710.1623389/frai.1622025.1627773. https://doi.org/doi:10.3389/frai.2025.1627773

HERA Digital. (2026). *Hera's Free & Open Source Technology including new AI-Powered Tools*. https://heradigitalhealth.org/ourtech/

HRW. (2021). *UN Shared Rohingya Data Without Informed Consent*. https://www.hrw.org/news/2021/06/15/un-shared-rohingya-data-without-informed-consent

Hunt, K., Narayanan, A., & Zhuang, J. (2022). Blockchain in humanitarian operations management: A review of research and practice. *Socio-Economic Planning Sciences*, *80*, 101175. https://doi.org/https://doi.org/10.1016/j.seps.2021.101175

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature machine intelligence*, *1*(9), 389–399.

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020, July). The State and Fate of Linguistic Diversity and Inclusion in the NLP World. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* Online.

Kathy Peach, A. B. (2022). Resilience, Peace and Security: Local communities should be part of designing humanitarian AI. https://www.weforum.org/stories/2022/10/local-communities-should-be-part-of-designing-humanitarian-ai/

Lidén, K. (2026). Ethics and the governance of digital data in humanitarian action. *Disasters*, *50*(1), e70018.

Lythreatis, S., Acikgoz, F., & Yassine, N. (2026). Artificial intelligence in humanitarian aid: A review and future research agenda. *Technovation*, *151*, 103415. https://doi.org/https://doi.org/10.1016/j.technovation.2025.103415

Mittelstadt, B. D. (2019). AI Ethics - Too Principled to Fail? *ArXiv*, *abs/1906.06668*.

Mshelia, Y., Malgwi, B., Erike, A., Zirra, A., & Olaye, E. (2025). The Humanitarian- AI Interface: Field Perspectives to Innovative Emergency Preparedness and Response. In A. A. Ajibesin & N. R. Vajjhala (Eds.), *AI for Humanitarianism: Fostering Social Change Through Emerging Technologies* (1st ed., pp. 242). CRC Press.

Nearing, G., Cohen, D., Dube, V., Gauch, M., Gilon, O., Harrigan, S., Hassidim, A., Klotz, D., Kratzert, F., Metzger, A., Nevo, S., Pappenberger, F., Prudhomme, C., Shalev, G., Shenzis, S., Tekalign, T. Y., Weitzner, D., & Matias, Y. (2024). Global prediction of extreme floods in ungauged watersheds. *Nature*, *627*(8004), 559–563. https://doi.org/10.1038/s41586-024-07145-1

Network, C. (2022). *CALP Network Annual Report*. https://www.calpnetwork.org/publication/the-calp-network-annual-report-2021-22/

O'Connor, S., & Liu, H. (2023). Gender bias perpetuation and mitigation in AI technologies: challenges and opportunities. *AI & society*, 1–13. https://doi.org/10.1007/s00146-023-01675-4

Ortega-Bolaños, R., Bernal-Salcedo, J., Germán Ortiz, M., Galeano Sarmiento, J., Ruz, G. A., & Tabares-Soto, R. (2024). Applying the ethics of AI: a systematic review of tools for developing and assessing AI-based systems. *The Artificial intelligence review*, *57*(5), 110. https://doi.org/10.1007/s10462-024-10740-3

Parkes, P., Pillay, T. D., Bdaiwi, Y., Simpson, R., Almoshmosh, N., Murad, L., & Abbara, A. (2022).

Telemedicine interventions in six conflict-affected countries in the WHO Eastern Mediterranean region: a systematic review. *Confl Health*, *16*(1), 64. https://doi.org/10.1186/s13031-022-00493-7

Paton-Romero, J. D., Vinuesa, R., Jaccheri, L., & Baldassarre, M. T. (2022). State of Gender Equality in and by Artificial Intelligence. *IADIS International Journal on Computer Science and Information Systems*, *17*(2), 31–48.

Pizzi, M., Romanoff, M., & Engelhardt, T. (2020). AI for humanitarian action: Human rights and ethics. *International Review of the Red Cross*, *102*(913), 145–180. https://doi.org/10.1017/S1816383121000011

Ramiah, D. (2024). 5 ways AI can help crisis response around the world. https://www.undp.org/5-ways-ai-can-help-crisis-response-around-world

Salio, P., Stoll, L., D'Amen, D., De Coning, E., Msemo, H., Schmid, F., Agersten, S., Sivle, A., Simon, A., & Chasco, M. J. (2025). Nowcasting and Early Warning Systems across WMO Regions Associations: A Pre-EW4All Assessment. *Bulletin of the American Meteorological Society*, *106*(12), E2490–E2508.

SAPEA. (2025). *Artificial intelligence has potential to improve emergency and crisis management, but struggles to interpret complex situations - experts tell the European Commission*. https://research-and-innovation.ec.europa.eu/news/all-research-and-innovation-news/artificial-intelligence-has-potential-improve-emergency-and-crisis-management-struggles-interpret-2025-12-11_en

Scheuerman, M. K., Pape, M., & Hanna, A. (2021). Auto-essentialization: Gender in automated facial analysis as extended colonial project. *Big Data & Society*, *8*(2), 20539517211053712. https://doi.org/10.1177/20539517211053712

Spencer, S. W., & Masboungi, C. (2025). Enabling access or automating empathy? Using chatbots to support GBV survivors in conflicts and humanitarian emergencies. *International Review of the Red Cross*, *107*(930), 1309–1350. https://doi.org/10.1017/S1816383125100763

Sperling, M., & Schryen, G. (2022). Decision support for disaster relief: Coordinating spontaneous volunteers. *European journal of operational research*, *299*(2), 690–705. https://doi.org/https://doi.org/10.1016/j.ejor.2021.08.022

Su, A. (2025). Gender equality, AI, and the future of human rights. *University of Toronto Law Journal*, *75*(Supplement 1), 202–217. https://doi.org/10.3138/utlj-2025-0034

Sürmeli, A. (2024, 14). August 2024: Digital Tools Transforming Humanitarian Aid. https://publichealth.jhu.edu/center-for-global-digital-health-innovation/august-2024-digital-tools-transforming-humanitarian-aid

Teodoro, E., Guillén, A., Gottschalk, T., Pichierri, F., Tschalaer, M., Xanthaki, A., Boland, C., & Macías, M. (2022). D2. 5 Report on the Legal, Societal Impact and Ethical Monitoring of the ITFLOWS.

UN. (2025). *Internal Audit Report of the Office of Internal Oversight Service for the period from 1 July 2024 to 30 June 2025*. https://documents.un.org/doc/undoc/gen/g25/122/94/pdf/g2512294.pdf

UNHCR. (2023). Keeping UNHCR's Biometrics System up to date. https://www.unhcr.org/blogs/keeping-unhcrs-biometrics-system-up-to-date/

UNHCR. (2025). UNHCR Blogs. https://www.unhcr.org/blogs/category/digital-identity-and-registration/

UNHCR. (2026). *Guidance on Registration and Identity Management*. https://www.unhcr.org/registration-guidance/chapter3/registration-tools/

United Nations. (2024). *Governing AI for Humanity*. United Nations High-level Advisory Body on Artificial Intelligence.

UNOCHA. (2024). *Philippines Anticipatory Action Framework for Cyclones*. https://www.unocha.org/publications/report/philippines/anticipatory-action-framework-philippines-2024#:~:text=The%20Framework%20outlines%20an%20approach,arranged%20financing%20(the%20Money).

UNOSAT. (2026). *From Image to Impact*. https://unosat.org/

Van den Homberg, M. J. C., Gevaert, C. M., & Georgiadou, Y. (2020). The Changing Face of Accountability in Humanitarianism: Using Artificial Intelligence for Anticipatory Action. Politics and Governance, *8*(4), 456–467. https://doi.org/10.17645/pag.v8i4.3158

van Maanen, G. (2022). AI Ethics, Ethics Washing, and the Need to Politicize Data Ethics. *Digital Society*, *1*(2), 9. https://doi.org/10.1007/s44206-022-00013-3

Wang, Y., Han, Y., Luo, A., Xu, S., Chen, J., & Liu, W. (2024). Site selection and prediction of urban emergency shelter based on VGAE-RF model. *Scientific Reports*, *14*(1), 14368. https://doi.org/10.1038/s41598-024-

64031-6

Weber de Melo, W., Iglesias, I., & Pinho, J. (2025). Early warning system for floods at estuarine areas: Combining artificial intelligence with process-based models. *Natural Hazards*, *121*(4), 4615–4638.

WFP. (2026). *WFP SCOPE: Registering Households Online*. https://usermanual.scope.wfp.org/scope-cards/content/common_topics/registration/registering_households_online.htm

WMO, U., AU, ITU, IFRC. (2023). *Multi-hazard Early Warning for All Action Plan for Africa (2023-2027)*. https://hdl.handle.net/10568/135516

Yu, X., Hu, X., Song, Y., Xu, S., Li, X., Song, X., Fan, X., & Wang, F. (2024). Intelligent assessment of building damage of 2023 Turkey-Syria Earthquake by multiple remote sensing approaches. *npj Natural Hazards*, *1*. https://doi.org/10.1038/s44304-024-00003-0

Ziemke, J., Jayamaha, B., & Jahn, M. M. (2021). Crisis Mapping and Crowdsourcing in Complex Emergencies. In: Oxford University Press.