Global Questions: An Interdisciplinary Review
INTERDISCIPLINARY PERSPECTIVE | COMMISSIONED SYNTHESIS

Trustworthy AI for Disaster Anticipation and Humanitarian Response

Nicholas Hall - Corresponding author

Abstract

This interdisciplinary perspective examines trustworthy artificial intelligence for disaster anticipation and humanitarian response. The organizing question is how predictive tools can improve timeliness while preserving local knowledge, accountability, and do-no-harm practice. Ten related scholarly sources are synthesized through a decision-centered framework spanning problem definition, mechanism, measurement, evaluation, implementation, and governance. The review does not invent experiments, pooled estimates, or unreported quantitative results. It instead evaluates the strength and transferability of the available evidence, with particular attention to evaluating prediction independently from whether warnings are trusted and actionable. The resulting framework links technical or empirical performance to explicit use conditions and identifies tests that should precede wider adoption in multi-hazard early warning and humanitarian logistics.

Keywords: disaster risk, humanitarian response, early warning, responsible AI, global resilience

Synthesis lens	Principal question
Mechanism	the chain from sensing and forecasting through warning communication, resource allocation, and field action
Evaluation	lead time, false alarms, missed events, geographic bias, usability, and consequences for affected communities
Primary risk	evaluating prediction independently from whether warnings are trusted and actionable

1. Introduction

Research on trustworthy artificial intelligence for disaster anticipation and humanitarian response sits at the boundary between a promising component and a consequential decision. The core question is how predictive tools can improve timeliness while preserving local knowledge, accountability, and do-no-harm practice. Answering it requires more than assembling favorable results. It requires a chain of evidence that makes the problem boundary, mechanism, measurement, comparator, uncertainty, and accountable decision visible to the reader.

This article presents a structured narrative review of ten related publications. Sources were selected for topical relevance, traceable publication metadata, and complementary coverage of technical, empirical, and governance questions. No meta-analytic pooling is attempted because the included studies differ in designs, populations, system boundaries, and outcomes. The purpose is decision-oriented synthesis: to identify what each source can support, where transfer remains uncertain, and which evidence would justify the next stage of use.

The central risk is evaluating prediction independently from whether warnings are trusted and actionable. A top-line metric or broad policy label can appear decisive while concealing the conditions that produced it. Accordingly, the synthesis separates observation from interpretation and implementation from validation. This separation is especially important when the same system creates benefits for one user or institution while transferring cost, risk, or administrative burden to another.

2. Evidence Base and Problem Boundary

A defensible review begins by defining the unit of analysis. For the present topic, the relevant boundary includes inputs, institutional or technical context, the mechanism producing an output, the person or system acting on that output, and the downstream outcome. Evidence falls outside the boundary when it measures only an intermediate proxy yet is used to claim an end-to-end benefit.

Linardos et al. (2022) addresses "Machine Learning in Disaster Management: Recent Developments in Methods and Applications." In this synthesis, that work is used as a source on problem definition within trustworthy artificial intelligence for disaster anticipation and humanitarian response. The connection is deliberately bounded: the cited study informs the evidence map, but it does not by itself establish readiness for multi-hazard early warning and humanitarian logistics. That distinction keeps the review close to the published record and prevents an adjacent result from being converted into a broader causal claim.

Ghaffarian et al. (2023) addresses "Explainable artificial intelligence in disaster risk management: Achievements and prospective futures." In this synthesis, that work is used as a source on conceptual boundary within trustworthy artificial intelligence for disaster anticipation and humanitarian response. The connection is deliberately bounded: the cited study informs the evidence map, but it does not by itself establish readiness for multi-hazard early warning and humanitarian logistics. That distinction keeps the review close to the published record and prevents an adjacent result from being converted into a broader causal claim.

Abid et al. (2021) addresses "Toward an Integrated Disaster Management Approach: How Artificial Intelligence Can Boost Disaster Management." In this synthesis, that work is used as a source on measurement within trustworthy artificial intelligence for disaster anticipation and humanitarian response. The connection is deliberately bounded: the cited study informs the evidence map, but it does not by itself establish readiness for multi-hazard early warning and humanitarian logistics. That distinction keeps the review close to the published record and prevents an adjacent result from being converted into a broader causal claim.

Together, these works provide complementary entry points, but their conclusions should not be flattened into a single ranking. Differences in data generation, setting, temporal horizon, and outcome definition may be substantively meaningful. A review should therefore preserve those differences and state where analogy, rather than direct replication, connects one domain to another.

3. Mechanism and System Design

The operative mechanism is the chain from sensing and forecasting through warning communication, resource allocation, and field action. This formulation discourages component-only reasoning. A model, platform, policy, assay, or infrastructure service acquires practical meaning only when its interfaces and use conditions are specified. The evidence chain should describe what information is available, what transformation occurs, which assumptions make that transformation credible, and who is authorized to act.

Aboualola et al. (2023) addresses "Edge Technologies for Disaster Management: A Survey of Social Media and Artificial Intelligence Integration." In this synthesis, that work is used as a source on system mechanism within trustworthy artificial intelligence for disaster anticipation and humanitarian response. The connection is deliberately bounded: the cited study informs the evidence map, but it does not by itself establish readiness for multi-hazard early warning and humanitarian logistics. That distinction keeps the review close to the published record and prevents an adjacent result from being converted into a broader causal claim.

Velev and Zlateva (2023) addresses "CHALLENGES OF ARTIFICIAL INTELLIGENCE APPLICATION FOR DISASTER RISK MANAGEMENT." In this synthesis, that work is used as a source on representation choice within trustworthy artificial intelligence for disaster anticipation and humanitarian response. The connection is deliberately bounded: the cited study informs the evidence map, but it does not by itself establish readiness for multi-hazard early warning and humanitarian logistics. That distinction keeps the review close to the published record and prevents an adjacent result from being converted into a broader causal claim.

Gupta et al. (2022) addresses "Artificial intelligence and cloud-based Collaborative Platforms for Managing Disaster, extreme weather and emergency operations." In this synthesis, that work is used as a source on failure analysis within trustworthy artificial intelligence for disaster anticipation and humanitarian response. The connection is deliberately bounded: the cited study informs the evidence map, but it does not by itself establish readiness for multi-hazard early warning and humanitarian logistics. That distinction keeps the review close to the published record and prevents an adjacent result from being converted into a broader causal claim.

These studies also show why modular claims require interface tests. A component may perform well while the complete pathway fails because inputs drift, users interpret outputs differently, recovery semantics are ambiguous, or institutional incentives change. Design documentation should therefore include not only the intended pathway but also foreseeable deviations, degraded modes, and conditions under which the system should stop or defer.

4. Evaluation and Uncertainty

Evaluation should center on lead time, false alarms, missed events, geographic bias, usability, and consequences for affected communities. Average performance remains useful, but it should be accompanied by distributions, error categories, relevant subgroups or operating regimes, and uncertainty at the point where a decision changes. Comparators must isolate the value of the proposed addition rather than compare a mature intervention with an intentionally weak baseline.

Mu et al. (2024) addresses "Making food systems more resilient to food safety risks by including artificial intelligence, big data, and internet of things into food safety early warning and emerging risk identification tools." In this synthesis, that work is used as a source on comparative evaluation within trustworthy artificial intelligence for disaster anticipation and humanitarian response. The connection is deliberately bounded: the cited study informs the evidence map, but it does not by itself establish readiness for multi-hazard early warning and humanitarian logistics. That distinction keeps the review close to the published record and prevents an adjacent result from being converted into a broader causal claim.

Kankanamge et al. (2021) addresses "Public perceptions on artificial intelligence driven disaster management: Evidence from Sydney, Melbourne and Brisbane." In this synthesis, that work is used as a source on uncertainty within trustworthy artificial intelligence for disaster anticipation and humanitarian response. The connection is deliberately bounded: the cited study informs the evidence map, but it does not by itself establish readiness for multi-hazard early warning and humanitarian logistics. That distinction keeps the review close to the published record and prevents an adjacent result from being converted into a broader causal claim.

External validation should introduce meaningful shifts in setting, time, population, workload, market structure, or environmental conditions. A nominally independent sample can still be too similar to test transfer. Conversely, a failed transfer is scientifically informative when the changed conditions are measured and the mechanism of failure is examined rather than hidden in an aggregate score.

5. Translation and Governance

Translation to multi-hazard early warning and humanitarian logistics depends on more than technical readiness. Decision rights, documentation, maintenance, monitoring, appeal, and recovery are part of the intervention. The governance requirement is participatory design, contestability, data protection, local authority, and post-event audit. Each control should be connected to a named failure mode and should identify an owner, evidence source, review interval, and escalation path.

Alam et al. (2019) addresses "Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of Hurricanes Harvey, Irma, and Maria." In this synthesis, that work is used as a source on implementation within trustworthy artificial intelligence for disaster anticipation and humanitarian response. The connection is deliberately bounded: the cited study informs the evidence map, but it does not by itself establish readiness for multi-hazard early warning and humanitarian logistics. That distinction keeps the review close to the published record and prevents an adjacent result from being converted into a broader causal claim.

Chamola et al. (2020) addresses "Disaster and Pandemic Management Using Machine Learning: A Survey." In this synthesis, that work is used as a source on governance within trustworthy artificial intelligence for disaster anticipation and humanitarian response. The connection is deliberately bounded: the cited study informs the evidence map, but it does not by itself establish readiness for multi-hazard early warning and humanitarian logistics. That distinction keeps the review close to the published record and prevents an adjacent result from being converted into a broader causal claim.

A credible implementation plan also defines sunset and revalidation conditions. New data, software updates, institutional restructuring, population change, or shifts in incentives can invalidate previously reasonable assumptions. Monitoring should therefore test the validity of the evidence chain, not merely whether a service remains online or a headline metric remains stable.

6. Research Agenda

Future studies should preregister or otherwise predefine the intended decision, principal outcomes, exclusions, and analysis plan when feasible. They should report missingness, sensitivity analyses, negative or null findings, and the cost of errors to differently situated stakeholders. Reproducible artifacts should include sufficient metadata to reconstruct data provenance, model or analytical versions, parameter choices, and the sequence from evidence to conclusion.

Cross-disciplinary work needs a shared object of explanation rather than a collection of adjacent vocabularies. For trustworthy artificial intelligence for disaster anticipation and humanitarian response, that shared object is the complete decision pathway. Technical, clinical, social, environmental, or economic expertise should each change the design or interpretation of the study. When evidence remains incomplete, the useful output is a bounded hypothesis, a transparent uncertainty statement, or a request for additional measurement.

7. Conclusion

The literature supports a disciplined approach to trustworthy artificial intelligence for disaster anticipation and humanitarian response: define the decision, preserve system boundaries, test consequential failure modes, connect uncertainty to action, and assign accountable control. The strongest conclusion is not that one method is universally superior. It is that credible use in multi-hazard early warning and humanitarian logistics requires an auditable link from source evidence to design choice, evaluation result, and governed decision.

References

Abid, S. K., Sulaiman, N., Chan, S. W., Nazir, U., Abid, M., Han, H., Ariza-Montes, A., & Vega-Muñoz, A. (2021). Toward an Integrated Disaster Management Approach: How Artificial Intelligence Can Boost Disaster Management. Sustainability, 13(22), 12560. https://doi.org/10.3390/su132212560

Aboualola, M., Abualsaud, K., Khattab, T., Zorba, N., & Hassanein, H. S. (2023). Edge Technologies for Disaster Management: A Survey of Social Media and Artificial Intelligence Integration. IEEE Access, 11, 73782-73802. https://doi.org/10.1109/access.2023.3293035

Alam, F., Ofli, F., & Imran, M. (2019). Descriptive and visual summaries of disaster events using artificial intelligence techniques: case studies of Hurricanes Harvey, Irma, and Maria. Behaviour and Information Technology, 39(3), 288-318. https://doi.org/10.1080/0144929x.2019.1610908

Chamola, V., Hassija, V., Gupta, S., Goyal, A., Guizani, M., & Sikdar, B. (2020). Disaster and Pandemic Management Using Machine Learning: A Survey. IEEE Internet of Things Journal, 8(21), 16047-16071. https://doi.org/10.1109/jiot.2020.3044966

Ghaffarian, S., Taghikhah, F. R., & Maier, H. R. (2023). Explainable artificial intelligence in disaster risk management: Achievements and prospective futures. International Journal of Disaster Risk Reduction, 98, 104123. https://doi.org/10.1016/j.ijdrr.2023.104123

Gupta, S., Modgil, S., Kumar, A., Sivarajah, U., & Irani, Z. (2022). Artificial intelligence and cloud-based Collaborative Platforms for Managing Disaster, extreme weather and emergency operations. International Journal of Production Economics, 254, 108642. https://doi.org/10.1016/j.ijpe.2022.108642

Kankanamge, N., Yigitcanlar, T., & Goonetilleke, A. (2021). Public perceptions on artificial intelligence driven disaster management: Evidence from Sydney, Melbourne and Brisbane. Telematics and Informatics, 65, 101729. https://doi.org/10.1016/j.tele.2021.101729

Linardos, V., Drakaki, M., Tzionas, P., & Karnavas, Y. (2022). Machine Learning in Disaster Management: Recent Developments in Methods and Applications. Machine Learning and Knowledge Extraction, 4(2), 446-473. https://doi.org/10.3390/make4020020

Mu, W., Kleter, G. A., Bouzembrak, Y., Dupouy, E., Frewer, L. J., Natour, F. N. R. A., & Marvin, H. J. P. (2024). Making food systems more resilient to food safety risks by including artificial intelligence, big data, and internet of things into food safety early warning and emerging risk identification tools. Comprehensive Reviews in Food Science and Food Safety, 23(1), e13296. https://doi.org/10.1111/1541-4337.13296

Velev, D., & Zlateva, P. (2023). CHALLENGES OF ARTIFICIAL INTELLIGENCE APPLICATION FOR DISASTER RISK MANAGEMENT. □The □international archives of the photogrammetry, remote sensing and spatial information sciences/International archives of the photogrammetry, remote sensing and spatial information sciences, XLVIII-M-1-2023, 387-394. https://doi.org/10.5194/isprs-archives-xlviii-m-1-2023-387-2023