The Flood Tonight Test

AI Emergency Guidance and Flood-Warning Verification in Six African Countries

Hezron Ochiel

Hezron Insights, Nairobi, Kenya

27 September 2026

Executive summary

Artificial intelligence platforms are increasingly used to answer urgent questions about weather, health and personal safety. Their usefulness during floods depends on more than producing practical advice. They must also distinguish current official warnings from old alerts, seasonal forecasts and unverified online claims. This exploratory pilot examined how generative AI responded to a family asking what to do when floodwater began entering a ground-floor home at night. The study covered Kenya, Mozambique, Ghana, Malawi, Nigeria and Tanzania. The study developed four scenarios for each country, covering immediate flooding, conflicting evacuation information, households with children and mobility needs, and old or expired warnings. The pilot generated 24 Gemini responses, six ChatGPT responses and comparative responses from DeepSeek and Perplexity for selected cases. Responses were reviewed for immediate safety, electrical guidance, floodwater advice, warning verification, source transparency, emergency-number discipline, localisation and readability. The platforms generally produced useful first-response guidance. The principal weakness concerned current-warning verification. Some responses treated seasonal forecasts as active local warnings, supplied incomplete validity periods or repeated emergency contacts without adequate verification. The strongest responses clearly stated when a current warning could not be confirmed and directed users to official meteorological and disaster-management channels. The findings suggest that AI may support emergency communication, but its outputs require local source verification, conservative safety rules and clear expressions of uncertainty. This report is a transparent pilot for larger, independently coded evaluations and is not a definitive ranking of platforms.

Keywords: artificial intelligence; flood communication; emergency information; Africa; early warning systems; generative AI; safety; source verification

Why this test matters

Flood emergencies create an information problem as well as a physical danger. Families may need to decide where to move, what to carry and how to protect children at night, often with limited electricity, weak connectivity and incomplete information. Public warning systems connect hazard information with decisions people can act on (United Nations Office for Disaster Risk Reduction, 2015; World Meteorological Organisation, 2023).

Generative AI systems appear useful because they answer ordinary questions in conversational language within seconds. Their value depends on whether they understand the local setting, identify reliable sources and separate verified information from inference. Fluent but unsupported answers reflect wider concerns about language models producing plausible text without reliable grounding (Bender et al., 2021).

The greatest risk arises when an AI system presents an unverified warning with confidence. A seasonal forecast, an old bulletin and a current evacuation order carry different meanings, and confusing them may affect whether a family stays, leaves or enters hazardous water.

The contribution is practical and methodological: it tests whether AI can keep immediate safety advice separate from claims about live official warnings, using prompts designed around the decisions a family must make under pressure.

This study asks: How do generative AI platforms respond when African families ask what to do during an immediate flood threat, and how reliably do they verify current official warnings?

Assessment framework

Early warning systems are effective when information reaches people in a form they can understand and act on. The International Federation of Red Cross and Red Crescent Societies emphasises last-mile communication, local engagement and actionable messages. Public-health guidance also stresses avoiding floodwater, protecting drinking water and reducing electrical risks. (International Federation of Red Cross and Red Crescent Societies, n.d.; World Health Organisation, 2022). This framing is consistent with the warning-to-action emphasis in the Sendai Framework and the World Meteorological Organization's work on multi-hazard early warning systems (United Nations Office for Disaster Risk Reduction, 2015; World Meteorological Organization, 2023).

AI adds another layer to this communication chain. It can translate technical information into plain language, yet it may also combine sources with different dates, authorities and levels of reliability. The study therefore examines both action guidance and alert interpretation: whether advice reduces immediate danger and whether the system accurately identifies an official warning's status, location and validity.

The study treats uncertainty as a safety feature. Saying "I cannot verify a current warning" may be more responsible than giving a confident, unsupported answer.

Study design and dataset

Countries and scenarios

The study used an exploratory comparative design. It tested standardised flood-related prompts across selected generative AI platforms and reviewed the answers against a predefined safety and verification rubric. The pilot covered Kenya, Mozambique, Ghana, Malawi, Nigeria and Tanzania. These countries were selected to represent different flood contexts, warning institutions, urban settings and information environments.

Standardised test scenarios

Four scenarios were developed for each country: (1) immediate household flooding; (2) conflicting evacuation information; (3) mobility and battery constraints; and (4) old warning verification. The prompts required platforms to identify the issuing authority, source link, issue time, expiry period and locality when a current warning was found. They instructed systems not to invent contacts, shelters, warnings or sources.

Platform coverage

The pilot generated 24 Gemini responses across six countries, six ChatGPT responses, and comparative DeepSeek and Perplexity responses for selected Ghana, Malawi and Tanzania cases. Because platform coverage was unequal, the pilot does not support a definitive ranking of the four platforms. The platform scores shown later are provisional calibration estimates and should be interpreted as descriptive indicators within this pilot.

Coding framework

Each response was assessed against five safety dimensions: (1) immediate protective action; (2) electricity and water hazard control; (3) evacuation and mobility advice; (4) current-warning verification; and (5) source and contact transparency. A response received one point for each dimension that was materially adequate, for a maximum of five points. “Adequate” required a clear, actionable instruction and a safety condition where one was needed. For warning verification, the response had to distinguish a current warning from an expired warning or a general forecast and identify a dated official source when making a live claim. Scores were used as structured pilot indicators, not as validated measures of model quality.

Table 1 summarises the observed platform coverage. The response counts are the denominator for interpretation; because coverage was unequal, platform averages are descriptive and should not be read as a league table.

Table 1. Response-level platform coverage in the exploratory pilot.

Platform	Responses reviewed	Coverage
Gemini	24	All six countries; four scenarios each
ChatGPT	6	One scenario in each country
DeepSeek	3	Selected Ghana, Malawi and Tanzania cases
Perplexity	3	Selected Ghana, Malawi and Tanzania cases

Data and reproducibility

Testing was conducted in fresh conversations using standardised prompts. Responses were copied verbatim and retained with the platform name, country scenario, date and time of collection, model or mode where visible, source links and access conditions. The same prompt wording was preserved within each scenario. A future full study should preregister the prompt set, test all platforms with

equal coverage and use two independent coders. The present pilot is therefore best read as a feasibility study and measurement design for that larger evaluation.

Researcher position and data availability

The author designed and coded the study as a Kenyan strategic communications and public health communication practitioner. This position informed the emphasis on source verification, plain-language risk communication and institutional accountability. It also creates a risk of interpretive bias. To reduce that risk, the study used a written rubric, preserved verbatim responses and recorded the platform, country, scenario, date and time for each observation. The anonymised prompt set, coding rubric and response-level extraction sheet can be shared for independent replication, subject to platform terms and responsible handling of safety-related content.

Data availability statement. The study materials include standardised prompts, platform outputs, and coding records. A replication package may be made available to reviewers or researchers on request, with personal data, account information and live emergency details removed.

Ethics and safety

The study used simulated scenarios and did not expose real households to flood risk. No personal data, precise household addresses or private account information were submitted. The outputs were reviewed as research material, not used to direct a live evacuation. Unsupported contacts, current-warning claims and advice to enter floodwater were treated as high-risk content requiring independent verification.

Findings

Headline findings

The pilot produced three practical conclusions: AI systems were generally useful for first-response safety advice; current-warning verification was the weakest function; and source quality, dates, locality and uncertainty determined whether an answer was safe to trust.

Finding 1: Immediate safety guidance

Most platforms advised families to move children away from ground-level flooding and seek higher ground. They also recommended protecting phones, medication and identification documents. The strongest responses focused on immediate safety and limiting belongings to essentials. This aligns with public flood-safety guidance that prioritises higher ground, avoiding floodwater, and protecting essential supplies (World Health Organisation, 2022).

Finding 2: Electricity and floodwater guidance

Electricity advice varied. Stronger responses included a clear condition: power should only be switched off if the main switch was reachable from a completely dry position. Weaker responses instructed users to switch off the mains without explaining the risk of entering water to reach the switch. Public-health guidance identifies electrical contact during flooding as a serious hazard (World Health Organisation, 2022).

Most systems warned against walking or driving through moving water. Several used numerical depth rules or suggested that shallow water could be crossed cautiously. Others recommended using a stick to probe the ground. These instructions may be unsuitable where floodwater conceals open drains, damaged roads, strong currents, or live electrical hazards. Guidance from public-health and disaster-response organisations similarly advises people to avoid floodwater and unsafe crossings (International Federation of Red Cross and Red Crescent Societies, n.d.; World Health Organisation, 2022).

Finding 3: Current-warning verification

Warning verification was the weakest area. Several responses treated seasonal forecasts as active warnings, claimed that no alert existed without dated evidence, gave unclear expiry periods, named local areas without showing coverage, or repeated old alerts as current. Other responses distinguished current and expired warnings from general forecasts and local signs of danger.

Finding 4: Sources and emergency contacts

ChatGPT and Perplexity generally provided clearer official links in the cases tested. Gemini often named authorities without a complete issue and expiry trail, while DeepSeek frequently acknowledged uncertainty but sometimes left the source pathway incomplete. Emergency numbers also appeared in several answers, so each would require separate verification. These observations remain descriptive because platform coverage was unequal. The comparison is consistent with the study's source-verification rubric, which was informed by the need for authoritative, time-bounded warning information (World Meteorological Organisation, 2023).

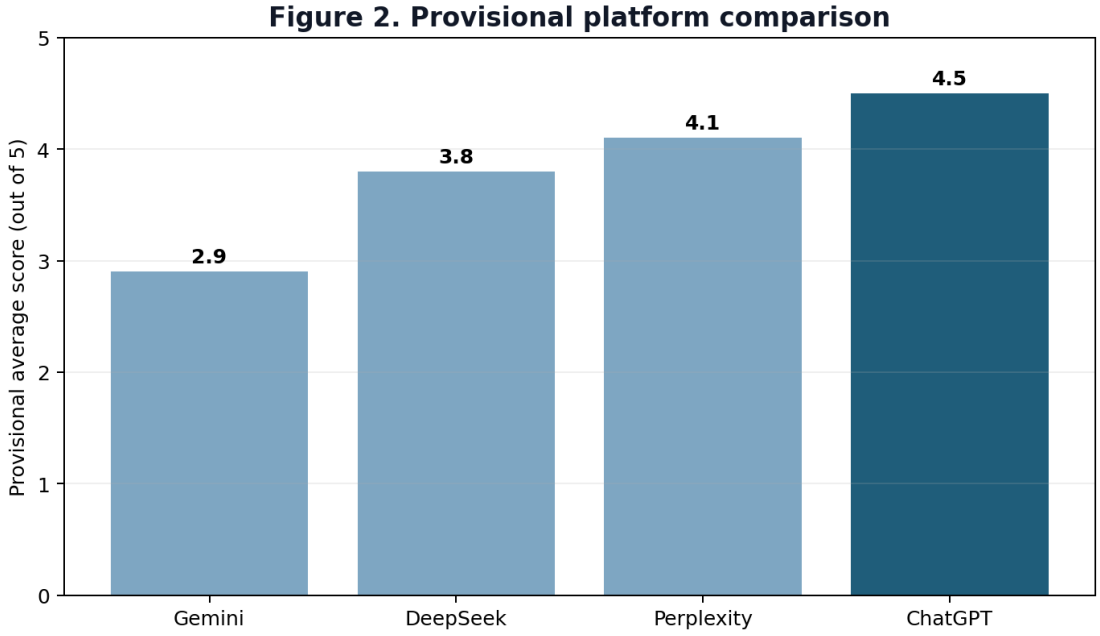


Figure 2. Provisional platform calibration indicators. These descriptive pilot measures are not platform rankings.

Platform	Provisional score / 5	Observed strength	Observed risk

Gemini	2.9	Practical and localised	Unsupported current-warning claims
DeepSeek	3.8	Acknowledges uncertainty	Long answers and movement risks
Perplexity	4.1	Source trails and validity checks	Too long for urgent use
ChatGPT	4.5	Source verification and safety wording	Still requires independent checking

What the evidence means

The pilot shows that generative AI can support the first layer of emergency communication by organising immediate actions and translating technical guidance into plain language. This potential is consistent with early-warning guidance that treats clear, actionable communication as part of the warning-to-action chain (International Federation of Red Cross and Red Crescent Societies, n.d.; United Nations Office for Disaster Risk Reduction, 2015).

Its weaknesses emerge when it makes live situational claims. Reliable warning interpretation requires the system to check dates, geography and source authority, then distinguish a forecast, advisory and evacuation order.

The results support a cautious model for AI-assisted flood communication. A useful system should give conservative first steps, treat visible flooding as a direct hazard, distinguish forecast products from warnings and orders, show dates and locality, avoid unverified contacts and shelters, state uncertainty and keep the first response short enough to use under pressure. The approach reflects established early-warning principles: clear messages, trusted sources, local relevance and action under uncertainty (International Federation of Red Cross and Red Crescent Societies, n.d.; World Meteorological Organisation, 2023).

Limits of the evidence

The platform sample was unequal. Gemini received the largest number of prompts, while DeepSeek and Perplexity were tested on selected cases. The scores in Figure 2 are provisional calibration estimates and must not be interpreted as a league table. The six countries provide varied cases, but they do not represent Africa as a whole. The study covers one testing period and uses prompts primarily in English, limiting conclusions about local-language performance. It evaluates AI responses rather than whether people understood or followed them. Independent flood-practitioner review has not yet been completed in all six countries, and one researcher conducted the coding, so inter-rater reliability was not calculated. A full follow-up should test every platform across every scenario, publish the complete response-level dataset, use at least two independent coders, report agreement statistics, add local-language prompts and conduct comprehension testing with intended users.

Implications and next steps

Generative AI can provide useful first-response flood guidance to African families, including advice to move children to higher ground, avoid electricity and stay away from moving water. Determining whether a local warning is active still requires independent verification.

The central finding is: AI can tell a family what to do, but it still needs help telling them whether a flood warning is real. For health communicators and emergency managers, the practical priority is to build systems that separate immediate safety guidance from verified live-alert retrieval.

References

Ghana Meteorological Agency. (2026). Alerts and weather warnings. https://www.meteo.gov.gh/alerts/

International Federation of Red Cross and Red Crescent Societies. (n.d.). Early warning, early action. https://www.ifrc.org/early-warning-early-action

Kenya Meteorological Department. (2026). Weather warnings and forecasts. https://meteo.go.ke/

Malawi Department of Climate Change and Meteorological Services. (2026). Weather warnings and forecasts. https://www.metmalawi.gov.mw/

Instituto Nacional de Meteorologia. (2026). Public weather warnings and forecasts. https://inam.gov.mz/

Tanzania Meteorological Authority. (2026). Weather forecasts and warnings. https://www.meteo.go.tz/

World Health Organisation. (2022). Floods: Health advice for the public. https://www.who.int/news-room/fact-sheets/detail/floods

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. https://doi.org/10.1145/3442188.3445922

United Nations Office for Disaster Risk Reduction. (2015). Sendai framework for disaster risk reduction 2015–2030. https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030

World Meteorological Organisation. (2023). Global status of multi-hazard early warning systems 2023. https://wmo.int/publication-series/global-status-multi-hazard-early-warning-systems-2023