The Impact of LLM-Generated Disaster Explanations on the Perceived Reliability and Clarity of Crisis News

by

A.R. Sao

S1083030

International Business Communication

Mentor: E.P. Sanders

Faculty of Arts

Radboud University

06-06-2025

Abstract

In times of crisis, clear and trustworthy communication is crucial. With the rise of large language models (LLMs) like ChatGPT, it is increasingly relevant to assess whether AI-generated explanations (rationales) for crisis-related news are perceived as reliable and clear as those written by humans. This study examined the perceived reliability and clarity of human- versus LLM-generated rationales for crisis tweets. A total of 114 participants took part in the experiment and were randomly assigned to read either a human- or a LLM-generated rationale. Participants rated the clarity and reliability of the rationale they read.

An independent samples *t*-test revealed no significant differences in perceived reliability or clarity between the two conditions. Human-generated rationales ($M = 13.7$, $SD = 3.5$) and LLM-generated rationales ($M = 14.1$, $SD = 3.0$) were rated similarly in reliability. Likewise, clarity ratings were nearly identical (human: $M = 9.9$, $SD = 2.4$; LLM: $M = 9.6$, $SD = 2.4$). In addition, a significant positive correlation was found between perceived reliability and clarity ($r = .63$, $p < .001$), suggesting that participants who found a rationale more reliable also perceived it as clearer.

These findings suggest that LLM-generated rationales can be just as effective as human-written ones in terms of clarity and trustworthiness, highlighting the potential role of AI in effective crisis communication.

Table of contents

Introduction

In today's digital landscape, social media platforms such as X (formerly known as Twitter), Instagram, and Facebook play a vital role in the dissemination of information. These platforms allow users to instantly share updates, photos, and commentary, which becomes especially important during crisis or disaster events. In these situations, timely and accurate information can guide emergency responses, enable situational awareness, and inform the public and humanitarian organizations. Individuals share details ranging from eyewitness accounts to urgent calls for help. The real-time nature of this information makes is a valuable resource for decision-makers, enabling quicker responses to emergencies such as natural disasters, public health crises, or terrorist attacks (Imran et al., 2020)

However, this abundance of social media data also presents challenges. The massive influx of posts during a disaster makes it difficult to filter relevant information from noise. Posts may include misinformation, personal commentary, or irrelevant content that hampers effective decision-making. Extracting useful insights requires fine-grained analysis that goes beyond simple keyword searches (Alam et al., 2021). Verifying the authenticity of information and summarizing it in a digestible and interpretable way are key obstacles. As a result, there has been a growing interest in using Artificial Intelligence (AI), particularly large language models (LLMs), to process social media content more efficiently.

LLMs are advanced models trained on vast datasets of text, capable of generating coherent, human-like language. These models can perform a range of tasks in natural language processing (NLP), such as classification, translation, summarization, and question answering. One of their key strengths is their ability to operate with minimal task-specific training while producing high-quality output. In the context of disaster management, LLMs have the potential to support efforts by automatically identifying relevant posts, summarizing complex information, and generating insights that can aid emergency decision-making (Zhang et al., 2024). LLMs have also been used in disaster classification systems like Rweetminer, which leverages machine learning to detect tweets containing urgent help requests further illustrating the operational potential of automated rationales in emergency contexts (Ullah et al., 2021).

In the field of disaster informatics, social media monitoring has emerged as a key strategy to supplement traditional information channels. Platforms like X and Facebook offer real-time, user-generated data that reflect the immediate conditions and needs of affected populations. This information can be critical for facilitating aid and allocating resources.

However, the chaotic and unstructured nature of social media makes automated processing both necessary and difficult. Extracting meaningful insights from noisy and high-volume data streams requires intelligent systems capable of distinguishing relevant crisis-related content from irrelevant or misleading posts (Alam et al., 2021).

LLMs offer a promising solutions. By leveraging deep neural networks trained on extensive language corpora, they can support disaster information through tasks such as classification of emergency-related posts, summarization of crisis developments, and generation of actionable insights. Their ability to analyze across thousands of messages in real time makes them valuable for humanitarian organizations. Besides that, the usefulness of LLMs is not only defined by their technical performance, but also by how their outputs are interpreted and trusted by human users (Zhang et al., 2024).

Zahra et al. (2019) additionally show how direct eyewitness messages differ from direct ones in structure and tone, direct messages often use sensory language like “I see smoke, “ whereas indirect posts tend to contain emotional responses or general reflections. Recognizing these linguistic cues is essential for training models that can effectively classify and prioritize content in emergencies.

Despite their promises, LLMs are not without limitations. A major concern is their black-box nature. These models often produce results without clear explanations of how or why those outputs were generated. This lack of transparency poses a serious challenge, especially in high-stakes contexts like emergency response, where trust and accountability are critical (Alam et al., 2021). Users need to understand the rationale behind AI-generated outputs, particularly when lives and resources are at stake.

To improve the interpretability of LLMs, researchers have explored the use of rationales – natural language explanations that accompany AI predictions. Rationales are intended to clarify why a model made a specific decision, making it more understandable to users (Gurrapu et al., 2023). When used correctly, rationales can bridge the gap between automated systems and human decision-makers, providing transparency and improving trust. The hope is that rationales will allow users to better assess the quality and accuracy of AI-generated content.

However, rationales generated by LLMs are not always reliable or truthful. For example such as Guo et al. (2023) show that LLMs sometimes generate plausible but incorrect explanations. This can undermine user trust, particularly if the rationale is misleading. Research by Mishra et al, (2024) emphasizes the importance of faithfulness – ensuring that rationales reflect the model’s actual reasoning process. A rationale that is

coherent but fabricated may appear trustworthy while failing to provide genuine insight into the model's decision-making. This tension highlight a critical concern: how rationales affect user perceptions of clarity and reliability.

Emprical studies have further examined how people respond to LLM-generated rationales. Leinonen et al. (2023) found that students, non-expert users, often preferred rationales generated by LLMs over those created by human annotators. However, this preference diminished when the rationales were paired with incorrect predictions. These findings raises concerns about the reliability of LLM-generated rationales in critical contexts where misinformation can have real-world consequences.

Recent research has begun to compare human and LLM rationality across decision-making. Alsagheer et al. (2024) examined how well LLMs align with human rationality, revealing both strengths and weaknesses in AI-generated reasoning. Their findings suggest that while LLMs has a broad and powerful knowledge base, their reasoning may diverge from human logic, especially in nuanced or emotionally charged contexts. Furthermore, irrational human feedback can impact the effectiveness of LLMs trained via reinforcement learning from human feedback (RLHF). These results underscore the need for more research into how rationality is perceived and measured in both human and machine-generated content.

There are also practical implications for using LLM-generated content in high-stakes settings. Rashid et al. (2024) demonstrated that LLM-generated educational content was more readable and sometimes even rated as higher quality than human-written texts. While these results are promising, they also raise concerns about over-reliance on AI without critical evaluation. Similarly, studies on creative writing by Köbis and Mossink (2020) found that people often could not distinguish between human-written and AI-generated poetry, suggesting that LLMs are increasingly effective in mimicking human language – but not necessarily human reasoning. Besides that, Sardinha (2023) argues that despite superficial similarities, LLM-generated texts still differ significantly from human-authored content at the structural and linguistic levels. These findings emphasize the complexity of replicating genuine human-like communication and rationality.

As LLMs become more integrated into crisis response systems, explainability becomes important. Rationales, textual justifications accompanying model outputs, play a key role in making LLM decisions understandable. They help bridge the gape between machine logic and human reasoning. Previous work, including studies Gurrapu et al. (2023) and also Mishra et al. (2024), highlights that users are more likely to trust model predictions

when supported by clear and faithful rationales. Yet, the accuracy of these explanations remains a concern: many rationales are generated through prompting techniques that may not precisely reflect the internal logic of the model (Fayyaz et al., 2024).

While existing studies have shown that LLMs are capable of generating rationales and categorizing disaster-related content, there is still limited understanding of how these explanations are perceived in terms of reliability and clarity. Much of the current literature focuses on the technical performance of LLMs, such as their classification accuracy. However, little is known about whether these AI-generated rationales genuinely support public trust in crisis communication or how their clarity may shape perceived reliability.

This study aims to address that gap by examining how people perceive the reliability and clarity of LLM-generated versus human-generated rationales in crisis news. As social media continues to be a dominant platform for the dissemination of real time disaster information, it is important to understand whether the type of rationale influences how credible and understandable that information is perceived to be. Insights from this research could support the development of more transparent and user-aligned LLMs, particularly in high-stakes contexts.

Thus, the research question is as followed: “How do LLM-generated disaster rationales affect the perceived reliability and clarity of crisis news compared to human rationales?” It is also hypothesized that participants will perceive human-generated rationales as more reliable and clearer than LLM-generated rationales.

Methodology

Materials

For this experimental study, the type of rationale source was manipulated to examine its effect on the perceived reliability and clarity of tweets. The independent variable is the type of rationale, with two levels: LLM-generated and human-generated.

This variable is operationalized by presenting participants with crisis-related tweets that include rationales explaining the information in the tweet. The key difference between the two versions of the stimulus material is the source of these rationales. One group of participants will see tweets with rationales written by humans, while another group will see the same tweets with rationales generated by Llama 3.2, Meta's large language model (LLM) designed for natural language processing tasks such as text generation, classification, and reasoning.

The tweets used in the study were collected from X and focus on a Philippine Flood Event in 2024. A total of 1200 tweets were divided among researchers in the Bachelor thesis group for annotation based on a predefined protocol. These human annotations served as the human-generated rationales and classifications. To avoid contamination, the LLM-generated rationales were only produced after all human annotations were completed, and were provided by the supervisor.

The classification labels used in this study were based on the humAID dataset (Alam et al., 2021), which contains disaster-related tweets. From the full set of available labels, five were selected: *Caution and advice, Infrastructure and utility damage, Injured or dead people, Rescue, volunteering, or donation effort, and Sympathy and support.* These categories were chosen because they appeared frequently across different disaster events and were relevant to the topic of this study.

From the full dataset, a subset of tweets was selected where the human and LLM assigned the same classification label, but gave different rationales. After looking at the annotations, it was noted that the difference between them were small. It was mostly a difference of a few words annotated or not. The different rationales helped isolate the effect of the rationale's source. A balance selection was then made, consisting of 50 tweets, with 10 tweets per classification label, to ensure equal distribution across all categories.

To assess the consistency between annotations provided by the human coders and those generated by the LLM, intercoder reliability was evaluated using Cohen's Kappa. The calculated Cohen's Kappa, for the tokens annotated in 120 tweets, indicated a fair agreement

for the rationales annotated by two human coders ($\kappa = 0.386$, $p < .001$). This suggests some overlap in what was considered a rational justification, but also a meaningful divergence, especially in what each annotator highlighted as relevant. The strongest agreement occurred when both coders excluded tokens from rationale, while overlap in positively selected tokens was more limited. This was also the case when comparing the first human-annotated with the LLM-annotated rationales. The calculated Cohen's Kappa between these was ($\kappa = 0.296$, $p < .001$). This also indicated a fair agreement.

To further evaluate the comparability between rationale sources, a surface level review of the content revealed that although the classification labels were the same, the LLM and human rationales often differed in phrasing and lexical emphasis. Only 86 rationales were exactly the same, and 820 shared the same classification label but had different rationales.

The selected tweets were organized into 20 tweets blocks, each containing 5 tweets and featuring two different classification labels per block. Ten blocks contained human-generated rationales and ten blocks contained LLM-generated rationales. Each participant was randomly assigned to one tweet block.

The stimulus materials is provided in the appendix.

Subjects

The participants in this study were from a general public pool. There was no strict selection criteria regarding age, gender, or educational background, anyone could participate. However, because the study was conducted in English, the participants needed to understand English. The survey was primarily distributed in English, which implicitly communicated the language requirement. In cases where the invitation was sent in Dutch, it was explicitly mentioned that the survey would be in English.

A total of 114 participants (but 220 people started the survey) took part in this study, with 57 randomly assigned to each condition (LLM-generated rationale or human-generated rationale).

Participants who were shown human-generated rationales were not significantly affected more likely to identify with a particular gender than participants who viewed LLM-generated rationales. Specifically, 57.9% of participants identified as female, 39.5% as male, 0.9% as non-binary, and 1.8% preferred not to disclose their gender. A Chi-square test indicated that the distribution of gender across the two rationale type conditions did not

significantly differ, χ^2 (3, N = 114) = 4.10, p = .251. Gender, therefore, did not contribute to a significant relation between rationale type and participant demographics.

An independent samples *t*-test was conducted to examine the participants' age and the type of rationales. The average age of participants was 26.6 (SD = 8.5), with the youngest being 18 years old and the oldest participant 55 years. The average age of participants who evaluated human-generated rationales had a mean age of 35.4 (SD = 8.3), whereas those who evaluated LLM-generated rationales had a mean age of 26.8 (SD = 8.8). The difference in age, -0.40, 95% CI [-3.58, 2.79], was not significantly different (t (110.84) = -.25, p = .81). Equal variances were not assumed due to one participant preferring not to disclose their age.

For the education background, participants reported various educational backgrounds. The most common educational level was bachelor's degree (N = 63, 55.3%), followed by high school diploma or equivalent of one (N = 26, 22.8%), and master's degree (N = 18, 15.8%), and other qualifications or possibly no qualifications (N = 7, 6.1%). A Chi-square test showed no significant relation between participants 'educational level and the type of rationale they were exposed to, χ^2(3, N = 114) = 1.81, p = .614. This indicates that the distribution of educational background was comparable across both experimental conditions.

The majority of participant in this study were Dutch (n = 54, 47.4%), followed by Chinese (n = 28, 24.6%). These two nationalities were more prominent in this study, other nationalities were represented by smaller numbers. Nationality was not considered an inclusion criterion and was collected for descriptive purposes. A chi-square test showed no significant difference in nationality distribution between the two conditions, χ^2 (19, N = 114) = 19.41, p = .43. This indicates that nationality was similarly distributed across both conditions.

Across all measured demographics variables, there were no significant differences between participants who saw human-generated rationales and those who saw LLM-generated rationales. The results showed that the random sampling was successful and the two different groups were comparable on the demographic variables. Therefore, any observed differences in the dependent variables are more likely due to the manipulation of the type of rationale (human vs. LLM), and not due to demographic factors.

Design

This study used a between-subjects design, meaning each participant was only exposed to one level of the independent variable. Participants were randomly assigned to one of two groups. One group read tweets that included human-generated rationales, while the other group read

tweets with LLM-generated rationales. In total, 57 participants were assigned to each condition, making the distribution equal across the groups.

After reading the tweets, participants were asked to rate the reliability and clarity of the rationale in the tweet. Because each participants only evaluated one type of rationale, this setup supports a between-subjects design. This approach allowed for a direct comparison between the two conditions and helped assess whether the type of rationale (human vs. LLM) influenced how people perceived the messages.

Instruments

The instrument used in this research is a questionnaire with several items to measure each dependent variable. The dependent variables were perceived reliability and perceived clarity on the type of generated rationale.

Perceived reliability was measured with three statements anchored by a seven-point Likert scale (strongly disagree – strongly agree), each introduced with: "To what extent do you agree/disagree with the following statements based on the block of tweets above:"

1. I trust the rationales given to justify the classification
2. The rationales demonstrate a high level of credibility
3. I am confident that the rationale in this tweet is accurate

The reliability of 'perceived reliability' comprising three items was acceptable: $\alpha = .78$'. Consequently, the mean of all three items was used to calculate the compound variable 'perceived reliability', which was used in the further analyses.

Perceived clarity was also measured with three statements anchored by a seven-point Likert scale (strongly disagree - strongly agree). Preceded by the same introduction: "to what extent do you agree/disagree with the following statements based on the block of tweets above:"

1. The rationale clearly explained the situation in the tweet
2. The rationale helped me understand the main message of the tweet
3. The rationale in this tweet was confusing

The third item was reverse-worded to reduce response biases such as acquiescence bias (Wong, 2003) and to assess whether the participants were paying attention. However, the reliability of 'perceived clarity', which initially included three items, was not acceptable: $\alpha = .47$. Further analysis indicated that the reverse-coded item significantly lowered the internal consistency. After examining the response patterns, it was revealed that many participants who agreed with the other two positively worded items, also agreed with the reverse-coded

item. This pattern is logically inconsistent. If participants found the rationale clear and helpful for understanding the situation, it is unlikely they would also genuinely perceive it as confusing. Therefore, the agreement with the reverse-coded item likely does not reflect a genuine judgment of low clarity. Instead, it may be the result of not paying attention, misunderstanding of the reversed phrasing, or survey fatigue. As a result, the item was excluded from the final scale to preserve validity and reliability. When removed, Cronbach's alpha increased to $\alpha = .62$. Although this is slightly below the ideal threshold, it is acceptable for a two-item scale, particularly since reliability coefficients tend to decrease with fewer items (Taber, 2017). A significant correlation between the two retained items ($p < .001$) supported their combined use, and their average was used to compute the compound variable 'perceived clarity' in further analyses.

Procedure

The experiment began with the annotation of tweets. In total 1200 tweets were manually annotated by researchers within the thesis group. To avoid any bias, LLM-rationales were added only after the human annotations were completed. From this annotated set, a final selection was made where the classifications from humans and AI matched, ensuring a balanced and reliable dataset for the experiment. This resulted in 820 tweets with matching classifications. Of these, only 86 tweets had completely identical rationales between the human- and LLM-annotated rationales. Since the objective of this research was to compare how participants perceive the reliability and clarity of different type of rationales, only tweets with matching classification labels but different rationales were included.

The distribution of matched classification labels (with different rationales) was as follows: 16 tweets for *Caution and Advice*, 185 for *Infrastructure and Utility Damage*, 340 for *Injured or Dead People* 146 for *Rescue, Volunteering, or Donation Effort*, and 49 for *Sympathy and Support.* From this dataset, a balanced subset of 50 tweets was randomly selected for the experiment to ensure an even representation for the experiment to ensure an even representation across all categories, which is already explain in the materials section.

Based on this dataset, an online survey was developed to measure the dependent variables of interest: perceived reliability and perceived clarity. The questionnaire was created collaboratively within the thesis group, and it included all necessary items to properly investigate the research questions.

The survey was distributed online through the personal networks of the researchers. Participants could take part on any device – smartphone, laptop, or tablet. At the start of the

survey, participants were given a brief explanation of how the survey worked and what a rationale is. However, to avoid influencing their responses, they were not told whether the rationales they would evaluate were human- or LLM-generated. Only in the final two questions were they asked whether they thought they has seen human- or AI-generated rationales, without revealing which one they actually got.

Participants were randomly assigned to one of two conditions: one group evaluated crisis-related tweets with human-generated rationales, and the other with LLM-generated rationales. After reading the tweets, they completed questions measuring dependent variables, for this research it was perceived reliability and clarity. All participants followed the same procedure, and the entire study was conducted individually online to ensure consistency.

Initially, the survey was estimated to take about 15 minutes to complete. However, the actual completion times varied widely. The time captured reflects the total duration between the moment a participant opened the survey and when they submitted it. This includes instances where participants may have paused or temporarily left the survey open. On average, it took participants 2 hours and 48 minutes to finish, with the minimum being under a minute and the maximum being almost a week. It is important to note thar this is not the net active time spent on the survey. This was partly due to the complexity of the task and the length of the instructions. Many participants reported that they found the introduction and instructions too long and skipped them, which led to confusion about how to proceed with the survey. As a result, many participants dropped out before completing it. Although around 220 people started the survey, only 114 complete responses were usable for analysis.

At the end of the survey, there was no formal briefing, as it was not considered necessary for the nature of this study. However, participants were thanked for their time and contribution.

Statistical treatment

An independent samples *t*-test was conducted to answer the research question: ‘How do LLM-generated disaster rationales affect the perceived reliability and clarity of crisis news compared to human-generated rationales?’ The hypotheses tested were:

- **H0**: There is no significant difference in perceived reliability and clarity of crisis news when comparing LLM-generated disaster rationales to human-generated rationales
- **H1**: There is a significant difference in perceived reliability and clarity of crisis news when comparing LLM-generated disaster rationales to human-generated rationales

Besides an independent samples *t*-test, a correlation test was also conducted between the variables of ‘perceived reliability’ and ‘perceived clarity’ to see if they were correlated.

Results

An independent samples *t*-test was used to test if there is a difference in perceived reliability and clarity of crisis news when comparing LLM-generated disaster rationales to human-generated rationales, with 'perceived reliability' and 'perceived clarity' as dependent variables and the type of rationale (human- vs LLM-generated) as the independent variable. The independent samples *t*-test on perceived reliability with type of rationale as between subject factor, showed no significant difference between the two conditions. This difference, -0.4, 95% CI [-1.61, 0.80], was not significant (t (112) = 0.66, p = .21). Thus, there was no main effect between perceived reliability on human-generated rationales (M = 13.7, SD = 3.5) and LLM-generated rationales (M = 14.1, SD = 3.0). For the perceived clarity, the independent samples *t*-test showed no significant differences between the type of rationale. This difference, 0.3, 95% CI [-0.57, 1.21], was not significant (t (112) = 45, p = .66). There was no main effected perceived clarity on human-generated rationales (M = 9.9, SD =2.4) and on LLM-generated rationales (M = 9.6, SD = 2.4). All these statistics can be found in Table 1.

Table 1

Comparison of Perceived Reliability and Clarity Scores between Human-Generated and LLM-Generated Rationales

	Type of Rationale	N	Mean	Std. Deviation	Std. Error Mean
PerceivedReliability	Human	57	13.7193	3.46284	.45866
	LLM	57	14.1228	3.01230	.39899
PerceivedClarity	Human	57	9.9298	2.40431	.31846
	LLM	57	9.6140	2.38862	.31638

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
PerceivedReliability	Equal variances assumed	1.628	.205	-.664	112	.508	-.40351	.60792	-1.60802	.80100
	Equal variances not assumed			-.664	109.892	.508	-.40351	.60792	-1.60827	.80126
PerceivedClarity	Equal variances assumed	.191	.663	.703	112	.483	.31579	.44890	-.57365	1.20523
	Equal variances not assumed			.703	111.995	.483	.31579	.44890	-.57365	1.20523

Other than an independent samples *t*-test, a correlation test was also performed. This was done to see if a rationale was perceived more reliable, this was correlated with the rationale being more clear. A significant positive relation was found between perceived reliability and perceived clarity on a rationale (r =.63, p < .001, N = 114). This indicates that the more reliable a rationale was perceived, the clearer it was also perceived to be. In other words, as participants rated a rationale as more trustworthy, they were also more likely to find it easier to understand. You can see these results in Table 2.

Table 2

Correlation Between Perceived Reliability and Perceived Clarity of Crisis News Rationales (High Scores Indicate Greater Perceived and Clarity)

		PerceivedReliability	PerceivedClarity
PerceivedReliability	Pearson Correlation	1	.625**
	Sig. (2-tailed)		.000
	N	114	114
PerceivedClarity	Pearson Correlation	.625**	1
	Sig. (2-tailed)	.000	
	N	114	114

**. Correlation is significant at the 0.01 level (2-tailed).

Conclusion and Discussion

The current research aimed to see if there is an effect of the type of rationale (human-generated or LLM-generated) on the perceived reliability and clarity of crisis-related information. The research question was as follows: "How do LLM-generated disaster rationales affect the perceived reliability and clarity of crisis news compared to human-generated rationales?"

The results showed no significant differences between the two types of rationales on either perceived reliability or perceived clarity. In other words, participants rated LLM-generated rationales as just as reliable and clear as those written by humans. This suggests that LLMs, despite their automated nature, can produce rationales that are perceived to be on par with those written human annotators, at least in terms of reliability and clarity within the context of disaster communication.

Additionally, a significant positive relation was found between perceived reliability and clarity. This suggests that when a rationale was seen as more reliable, it was also likely to be perceived as clearer. These findings indicate that the importance of clarity in shaping perceptions of credibility in shaping perceptions of credibility and supports the idea that the effectiveness of rationale may depend not only on its content but also on how it is presented.

Together, these outcomes show that LLMs may serve as a useful tool in generating explanations in crisis contexts, especially when human resources are limited. However, caution is still warranted, as perceptions do not necessarily reflect factual accuracy or reasoning transparency.

The present study investigated the impact of LLM-generated versus human-generates rationales on the perceived reliability and clarity of crisis-related content. While previous literature suggested that LLMs can closely approximate human reasoning in some contexts (Alsagheer et al., 2024), our findings revealed that no significant differences in perceived clarity or reliability between the types of rationales. These results suggest that, at least within the context of the current design, LLMs can produce rationales that are perceived as equally trustworthy and understandable as those written by humans. This aligns partially with previous findings from Leinonen et al. (2023), who reported that users preferred LLM rationales in specific contexts, though their study also revealed that declining trust when rationales accompanied incorrect outputs. The current study, by isolating correctly classified outputs, complements this by showing that, in the absence of classification errors, LLM-generated rationales are comparably effective.

This results also align with research from Köbis and Mossink (2020), who demonstrated that participants were often unable to distinguish between AI-generated and human-generated content. In a similar vein. Sardinha (2023) argued that while AI-generated text may differ from human-authored text at a structural level, it can still appear convincingly human to readers. The absence of difference in perceived clarity and reliability found here supports the idea that surface-level textual cues in rationales may be sufficient to create parity in perception, even if the underlying reasoning structure differs between LLMs and humans (Mishra et al., 2024).

However, our results differ somewhat from those of Fayyaz et al. (2024), who found that attribution-based rationales better aligned with human reasoning than prompting-based ones. While our study did not directly compare these methods, the lack of perceived difference in clarity or reliability may suggest that, at least in prompting-based formats, LLMs are able to approximate the explanatory value of human-authored rationales in the context of short social media posts.

Another interesting finding was the strong positive correlation between perceived reliability and clarity. This supports the notion that users view clear rationales as more trustworthy, or vice versa. The interdependence has important implications for AI work, reinforcing the importance to focus on both interpretability and clarity when developing explanatory systems for high-stakes settings like disaster response.

Despite these promising findings, several limitations needs to be acknowledged. First, the sample lacked diversity in terms of demographic background. Most participants were Dutch or Chinese university students recruited through personal networks, which limits the generalizability of the findings. Furthermore, the participants were mostly not native English speakers. Although the rationales and questions were written in English, and the participants had sufficient proficiency to complete the survey. Language comprehension may still have impacted their interpretation of the rationales.

A second limitation relates to the complexity of the survey. Participants reported that the survey was difficult to understand, particularly the instructions and the task of evaluating rationales. This may have contributed to inconsistent responses, especially on the reverse-coded item used to assess attention and comprehension. The contradictory response to the reverse-coded items suggest that some participants were either not paying sufficient attention or failed to understand the them. This phenomenon is not uncommon in a survey. Van

Sonderen et al. (2013) argued that reverse-worded items often reduce the reliability of scales and may confuse respondents, rather than serve as effective attention checks.

The third limitation is the length and structure of the survey led some participants to drop out. While approximately 220 individuals began the survey, only 114 completed it. This high dropout rate suggests that the survey may have been too long or too complex. Future studies should consider simplifying the instructions, reducing the number of questions, and pretesting the survey to enhance clarity and completion rates.

Another methodological consideration involves the stimuli themselves. The different types of rationales, human- vs LLM-generated, were based on a single disaster event, the Philippines flood in 2024. Besides that, only five of the original humAID classification labels were used. While this ensured consistency, it also limits the scope of applicability. Other disaster types, or more diverse sets of labels, may produce different findings.

Future studies could address several of these limitations. Firstly, to enhance generalizability, a broader and more diverse participant pool should be used. Including individuals from different linguistic and cultural backgrounds would offer a more a comprehensive understanding of how LLM-generated rationales are perceived. It could also be useful to group findings by age, as younger generations may be more familiar with AI tools and social media platforms, potentially influencing their perceptions of AI-generated content.

Secondly, future research could consider designing more effective measures of attention and comprehension that go beyond the use of reverse-coded items. Alternatives such as direct comprehension questions could be more reliable indicators of participant engagement.

Additionally, it would be valuable to examine how perceived reliability and clarity interact with actual accuracy. In this research, all rationales accompanied correctly classified tweets, but future experiments could systematically manipulate correctness to observe how users weigh clarity and reliability against outcome accuracy.

The type of crisis or disaster event may also influence how rationales are interpreted. Natural disasters, public health emergencies, and violent incidents may trigger different emotional and cognitive response. Therefore, replicating this study across diverse crisis types could help validate the findings in different contexts.

Lastly, incorporating qualitative feedback from participants could help understanding of how the interpret and evaluate rationales. Open-ended questions or interviews could

uncover subtle factors, such as emotional resonance or perceived intent, that influence judgements of clarity and trustworthiness.

In sum, this study contributes to the growing body of research on explainable AI by showing that, under controlled conditions, LLM-generated rationales can be perceived as equally reliable and clear as human-generated ones in the context of disaster-related tweets. This challenges concerns that LLMs inherently lack the trustworthiness or communicative nuance of human reasoning. However, methodological constraints, participants demographics, and survey design all suggest that further research is needed before generalizing these findings to broader populations or real-world settings.

By identifying the conditions under which LLM-generated rationales are effective, this study provides a foundation for future research aimed at enhancing the transparency and reliability of AI systems. Given the increasing use of LLMs in public communication and crisis response, ensuring that these systems are not only accurate but also comprehensible and trustworthy is of critical societal importance.

Reference list

Alam, F., Qazi, U., Imran, M., & Ofli, F. (2021). HumAID: Human-Annotated Disaster Incidents Data from Twitter with Deep Learning Benchmarks. *arXiv (Cornell University)*. https://doi.org/10.48550/arXiv.2104.03090

Alsagheer, D., Karanjai, R., Diallo, N., Shi, W., Lu. Y., Beyoun,, S., & Zhang, Q (2024). Comparing Rationality Between Large Language Models and Humans: Insights and Open Questions. *arXiv (Cornell University).* https://doi.org/10.48550/arxiv.2403.09798

Fayyaz, M., Yin, F., Sun. J., & Peng, N. (2024). Evaluating Human Alignment and Model Faithfulness of LLM Rationale. *arXiv (Cornell University).* *https://doi.org/10.48550/arxiv.2407.00219*

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding. Y., Yue, J., & Wu, Y. (2023). How close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection. *arXiv (Cornell University)*. https://doi.org/10.48550/arXiv.2301.07597

Gurrapu, S., Kulkarni, A., Huang, L., Lourentzou, I., Freeman, L., & Batarseh, F.A. (2023) Rationalization for Explainable NLP: A Survey. *arXiv (Cornell University).* https://doi.org/10.48550/arXiv.2301.08912

Imran, M.m Ofli, F., Caragea, D., & Torrralba, A. (2020). Using AI and Social Media Multimodal Content for Disaster Response and Management: Opportunities, Challenges, and Future Directions. *Information Processing & Management, 57*(5). https://doi.org/10.1016/j.ipm.2020.102261

Köbis, N., & Mossink, L. D. (2020) Artificial Intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior, 114*, 106553. https://doi.org/10.1016/j.chb.2020.106553

Leinonen, J., Denny, P., MacNeil, S., Sarsa, S., Bernstein, S., Kim, J., Tran, A., & Hellas, A. (2023). Comparing Code Explanations Created by Students and Large Language Models. *arXiv (Cornell University)*. https://doi.org/10.48550/arXiv.2304.03938

Mishra, A., Rahman, S., Mitra, K., Kim, H., & Hruschka, E. (2024). Characterizing Large Language Models as Rationalizers of Knowledge-intensive Tasks. *Findings of The Association For Computational Linguistics: ACL. 2022,* 8117-8139. https://doi.org/10.18653/v1/2024.findings-acl.484

Rashid, M. M., Atilgan, N., Dobres, J., Day, S., Penkova, V., Küçük, M., Clapp, S. R., &

Sawyer, B. D. (2024). Humanizing AI in Education: a Readibility Comparison of LLM and Human-Created Educational Content. *Proceedings of the Human Factors And Ergonomics Society Annual Meeting*. https://doi.org/10.1177/10711813241261689

Sardinha, T. B. (2023). AI-generated vs human-authored texts: A multiedimensional comparison. *Applied Corpus Linguistics, 4* (1), 100083. https://doi.org/10.1016/j.acorp.2023.100083

Taber, K. S. (2017). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education, 48* (6), 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Ullah, I., Khan, S., Imran, M., & Lee Y. (2021). RweetMiner: Automatic identification and categorization of help requests on twitter during disasters. *Expert Systems With Applications, 176*. https://doi.org/10.1016/j.eswa.2021.114787

Van Sonderen, E., Sanderman, R., & Coyne, J.C. (2013), Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS ONE, 8*(7), e68967. https://doi.org/10.1371/journal.pone.0068967

Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003. Do Reverse-Worded Items Confound Measures in Cross-Cultural Consumer Research? The Case of the Material Values Scale. *Journal of Consumer Research, 30* (1), 72 91. https://psycnet.apa.org/doi/10.1086/374697

Zahra, K., Imran, M., & Ostermann, F.O. (2019). Automatic identification of eyewitness messages on twitter during disasters. *Information Processing & Management, 57*(1). https://doi.org/10.1016/j.ipm.2019.102107

Appendix

Tweet blocks used in the surveys (H = human annotatted, L = LLM annotated):

Block 1 (H)

Tweet	Label
@Username Rescuers in Philippines are wading through chest-deep floodwaters to reach residents stranded by Tropical Storm Trami, which has claimed seven lives and forced thousands to evacuate as it approaches the east coast. **[URL Here]** #TropicalStormTrami #floods #Rainfall鈥	Injured or dead people
@Username Tropical Storm Trami struck the Philippines, killing 24, with flooding and landslides across Luzon. Rescue efforts continue as schools stay shut, and the death toll may rise. Emergency services are working overtime to manage the crisis. **[URL Here]**	Injured or dead people
@Username The Philippines canceled currency trading after most government agencies suspended operations, as Tropical Storm Trami tracked toward the main island bringing heavy rain that鈥 flooded some areas. > BSP: Thank you typhoon Trami (Kristine)? **[URL Here]**	Caution and advice
@Username Former Civil Service Commission Chair Karlo Nograles stressed the urgency of addressing the flooding woes of Davao City given the rising frequency of extreme weather events affecting the country. #PolitikoNews | @Username **[URL Here]**	Caution and advice
@Username PAG-BAHA SA DAVAO CITY SOLUSYUNAN NA! As Typhoon Kristine causes flooding in Bicol, former CSC Chair @Username emphasized the urgency of addressing Davao City鈥 flood problems amid increasing extreme weather events. READ MORE: **[URL Here]**	Caution and advice

Block 1 (L)

Tweet	Label

@Username Rescuers in Philippines are wading through chest-deep floodwaters to reach residents stranded by Tropical Storm Trami, which has claimed seven lives and forced thousands to evacuate as it approaches the east coast. **[URL Here]** #TropicalStormTrami #floods #Rainfall鈥	Injured or dead people
@Username Tropical Storm Trami struck the Philippines, killing 24, with flooding and landslides across Luzon. Rescue efforts continue as schools stay shut, and the death toll may rise. Emergency services are working overtime to manage the crisis. **[URL Here]**	Injured or dead people
@Username The Philippines canceled currency trading after most government agencies suspended operations, as Tropical Storm Trami tracked toward the main island bringing heavy rain that鈥 flooded some areas. > BSP: Thank you typhoon Trami (Kristine)? **[URL Here]**	Caution and advice
@Username Former Civil Service Commission Chair Karlo Nograles stressed the urgency of addressing the flooding woes of Davao City given the rising frequency of extreme weather events affecting the country. #PolitikoNews | @username **[URL Here]**	Caution and advice
@Username PAG-BAHA SA DAVAO CITY SOLUSYUNAN NA! As Typhoon Kristine causes flooding in Bicol, former CSC Chair @username emphasized the urgency of addressing Davao City鈥 flood problems amid increasing extreme weather events. READ MORE: **[URL Here]**	Caution and advice

Block 2 (H)

Tweet	**Label**
@Username Camarines Sur 2nd District Rep. LRay Villafuerte shut down rumors and denied viral photos on social media showing that he and his sons were in Siargao as Severe Tropical Storm Kristine flooded the Bicol Region. **[URL Here]**	Caution and advice

@Username The water level of La Mesa Dam in Quezon City is nearing its spilling level amid the non-stop rains on Thursday due to Severe Tropical Storm Kristine, according to PAGASA鈥檚 situationer. Click the photo to read more: **[URL Here]**	Caution and advice
@Username Bicol is currently facing its most severe flooding in the past years. WE NEED YOUR PRAYERS Please Pray for Bicol倮檎	Sympathy and support
@Username Some parts of my lapags tonight will go to my relatives in Bicol who are currently experiencing severe flooding. #PrayforBicol	Sympathy and support
@Username If you come across this tweet, please say a short prayer for the Philippines and everyone affected by the strong typhoon and massive flooding 倮椟倮檎 Prayer works. #KristinePH	Sympathy and support

Block 2 (L)

Tweet	**Label**
@Username Camarines Sur 2nd District Rep. LRay Villafuerte shut down rumors and denied viral photos on social media showing that he and his sons were in Siargao as Severe Tropical Storm Kristine flooded the Bicol Region. **[URL Here]**	Caution and advice
@Username The water level of La Mesa Dam in Quezon City is nearing its spilling level amid the non-stop rains on Thursday due to Severe Tropical Storm Kristine, according to PAGASA鈥檚 situationer. Click the photo to read more: **[URL Here]**	Caution and advice

@Username Bicol is currently facing its most severe flooding in the past years. WE NEED YOUR PRAYERS Please Pray for Bicol傈檹	Sympathy and support
@Username Some parts of my lapags tonight will go to my relatives in Bicol who are currently experiencing severe flooding. #PrayforBicol	Sympathy and support
@Username If you come across this tweet, please say a short prayer for the Philippines and everyone affected by the strong typhoon and massive flooding 傈椵傈檹 Prayer works. #KristinePH	Sympathy and support

Block 3 (Human)

Tweet	**Label**
@Username Sending prayers and strength to my PH moots and everyone affected by the floods in Bicol region. Please stay safe傈檹	Sympathy and support
@Username Hello, everyone! Let鈥 keep the people affected by #KristinePH in our thoughts and prayers, especially those in the Bicol region, which has been hit hard by continuous heavy rains and strong winds, causing severe flooding in many areas. My heart goes out to everyone affected by鈥 **[URL Here]**	Sympathy and support
@Username Former Vice President Leni Robredo aired an appeal for pump boats amid massive flooding in Naga City due to Tropical Storm Kristine. #AbogadoNews **[URL Here]**	Infrastructure and utility damage

@Username LOOK: During the height of Severe Tropical Storm Kristine on Tuesday night, floodwaters reached knee-high levels inside the Naga Metropolitan Cathedral. | Via Gabriela Baron, photos from Naga Metropolitan Cathedral #KristinePH #DailyTribune **[URL Here]**	Infrastructure and utility damage
@Username LOOK: The circumferential road in Barangay Sta. Barbara in the City of Ilagan, Isabela, is impassable due to above-head flooding brought by Severe Tropical Storm Kristine. #KristinePH made landfall in Divilacan, Isabela, at 12:30 a.m. today, Oct. 24. | 馃憖: City of Ilagan DRRMO - 鈥 https://t.co/3JwZgS1Exb	Infrastructure and utility damage

Block 3 (LLM)

Tweet	**Label**
@Username Sending prayers and strength to my PH moots and everyone affected by the floods in Bicol region. Please stay safe 馃檹	Sympathy and support
@Username Hello, everyone! Let鈥榮 keep the people affected by #KristinePH in our thoughts and prayers, especially those in the Bicol region, which has been hit hard by continuous heavy rains and strong winds, causing severe flooding in many areas. My heart goes out to everyone affected by鈥 https://t.co/xFlqIVUBq8	Sympathy and support
@Username Former Vice President Leni Robredo aired an appeal for pump boats amid massive flooding in Naga City due to Tropical Storm Kristine. #AbogadoNews **[URL Here]**	Infrastructure and utility damage

@Username LOOK: During the height of Severe Tropical Storm Kristine on Tuesday night, floodwaters reached knee-high levels inside the Naga Metropolitan Cathedral. | Via Gabriela Baron, photos from Naga Metropolitan Cathedral #KristinePH #DailyTribune **[URL Here]**	Infrastructure and utility damage
@Username LOOK: The circumferential road in Barangay Sta. Barbara in the City of Ilagan, Isabela, is impassable due to above-head flooding brought by Severe Tropical Storm Kristine. #KristinePH made landfall in Divilacan, Isabela, at 12:30 a.m. today, Oct. 24. | 馃摲: City of Ilagan DRRMO - 鈥 **[URL Here]**	Infrastructure and utility damage

Block 4 (Human)

Tweet	**Label**
@Username Tropical Storm Triggers Devastating Floods and Landslides in the Philippines #Philippines #floods #strom #landslides Read more: **[URL Here]**	Infrastructure and utility damage
@Username LOOK | NEWS UPDATE: A storm surge caused by #KristinePH floods areas near the Tondaligan Beach in Brgy. Bonuan Gueset, Dagupan City More details on GMA Regional TV One North Central Luzon, 5:10 p.m., on your local GMA channel. **[URL Here]**	Infrastructure and utility damage
@Username The local government unit of Naga City and former Vice President Leni Robredo on Wednesday urgently appealed for trucks, pump boats, and other resources to aid in rescuing flood victims affected by tropical storm Kristine.	Rescue, volunteering, or donation effort
@Username Not PBB related but rescues are desperately needed in Bicol Region. Many areas are already flooded. If you 鈥檓e in the affected areas, please follow local warnings and stay safe.	Rescue, volunteering, or donation effort

@Username Context: Robredo earlier joined calls for more rubber / pump boats to be lent to rescuers in Bicol after severe floods submerged houses in the area. The disaster risk and management council in Bicol has called for assistance as rescuers can no longer respond to the surge in鈥**[URL Here]**	Rescue, volunteering, or donation effort

Block 4 (LLM)

Tweet	**Label**
@Username Tropical Storm Triggers Devastating Floods and Landslides in the Philippines #Philippines #floods #strom #landslides Read more:**[URL Here]**	Infrastructure and utility damage
@Username LOOK | NEWS UPDATE: A storm surge caused by #KristinePH floods areas near the Tondaligan Beach in Brgy. Bonuan Gueset, Dagupan City More details on GMA Regional TV One North Central Luzon, 5:10 p.m., on your local GMA channel. **[URL Here]**	Infrastructure and utility damage
@Username The local government unit of Naga City and former Vice President Leni Robredo on Wednesday urgently appealed for trucks, pump boats, and other resources to aid in rescuing flood victims affected by tropical storm Kristine.	Rescue, volunteering, or donation effort
@Username Not PBB related but rescues are desperately needed in Bicol Region. Many areas are already flooded. If you 鈥檓e in the affected areas, please follow local warnings and stay safe.	Rescue, volunteering, or donation effort
@Username Context: Robredo earlier joined calls for more rubber / pump boats to be lent to rescuers in Bicol after severe floods submerged houses in the area. The disaster risk and management council in Bicol has called for assistance as rescuers can no longer respond to the surge in鈥 **[URL Here]**	Rescue, volunteering, or donation effort

Block 5 (Human)

Tweet	Label
@Username We are calling for donations for communities in Bicol affected by flooding and heavy rains brought on by Typhoon Kristine. No amount of help is too little. Join our donation drive! #Kristine #KristinePH #BrigadaKalikasan #CallForDonations	Rescue, volunteering, or donation effort
@Username AFP deploys HADR assets amid Tropical Storm Kristine CAMP AGUINALDO, Quezon City鈥攖he Armed Forces of the Philippines (AFP) has deployed multiple humanitarian assistance and disaster response (HADR) units across the country to rescue families affected by severe flooding caused by鈥 **[URL Here]**	Rescue, volunteering, or donation effort
@Username Philippine rescuers raced Thursday to reach residents still stranded by flooding in the hard-hit Bicol region after torrential rains from Tropical Storm Trami submerged villages and killed more than 20 people.	Injured or dead people
@Username Tropical Storm #Trami| Widespread #flooding and #landslides set off by a tropical storm in the Philippines left over 25 people dead, swept away cars and prompted authorities to rescue trapped villagers. **[URL Here]**	Injured or dead people
@Username At least 46 individuals were reported dead in the Philippines due to massive floods and landslides caused by #KristinePH (international name: Trami), data from the Office of Civil Defense shows. | @username Follow live updates on #KristinePH: **[URL Here]**	Injured or dead people

Block 5 (LLM)

Tweet	Label

@Username We are calling for donations for communities in Bicol affected by flooding and heavy rains brought on by Typhoon Kristine. No amount of help is too little. Join our donation drive! #Kristine #KristinePH #BrigadaKalikasan #CallForDonations	Rescue, volunteering, or donation effort
@Username AFP deploys HADR assets amid Tropical Storm Kristine CAMP AGUINALDO, Quezon City鈥攖he Armed Forces of the Philippines (AFP) has deployed multiple humanitarian assistance and disaster response (HADR) units across the country to rescue families affected by severe flooding caused by鈥 **[URL Here]**	Rescue, volunteering, or donation effort
@Username Philippine rescuers raced Thursday to reach residents still stranded by flooding in the hard-hit Bicol region after torrential rains from Tropical Storm Trami submerged villages and killed more than 20 people.	Injured or dead people
@Username Tropical Storm #Trami| Widespread #flooding and #landslides set off by a tropical storm in the Philippines left over 25 people dead, swept away cars and prompted authorities to rescue trapped villagers. **[URL Here]**	Injured or dead people
@Username At least 46 individuals were reported dead in the Philippines due to massive floods and landslides caused by #KristinePH (international name: Trami), data from the Office of Civil Defense shows. | @username Follow live updates on #KristinePH: **[URL Here]** https://twitter.com/inquirerdotnet/status/1849681008637116840	Injured or dead people

Block 6 (Human)

Tweet	**Label**
@Username 33 people were killed in landslides set off by Tropical Storm Trami in a province south of Manila, a Philippine police official said.	Injured or dead people

@Username The number of dead and missing in massive flooding and landslides wrought by Tropical Storm Trami in the Philippines has exceeded 100 and the president said Saturday that many areas remained isolated with people in need of rescue. **[URL Here]**	Injured or dead people
@Username Tropical Storm Trami triggers killer floods, landslides in Philippines and the death toll is mounting READ MORE: **[URL Here]** **[URL HERE]**	Injured or dead people
@Username Oh please pray to everyone who are affected of typhoon Kristine here in the Philippines 倮檹倮椢 hope everyone who needs help already rescue and no more hurt. The wind is so strong and heavy rain is here again 倮椢 please stop this already 倮檹倮橏 it's extreme alert in my place. **[URL Here]**	Sympathy and Support
@Username Dear God, we lift up the Bicol region to you during tropical storm #Kristine. The heavy rains have caused severe flooding, forcing thousands of families to evacuate. We pray for their safety and comfort. With several provinces under Signal #2, we ask for your divine protection釱 **[URL Here]**	Sympathy and Support

Block 6 (LLM)

Tweet	**Label**
@Username 33 people were killed in landslides set off by Tropical Storm Trami in a province south of Manila, a Philippine police official said.	Injured or dead people
@Username The number of dead and missing in massive flooding and landslides wrought by Tropical Storm Trami in the Philippines has exceeded 100 and the president said Saturday that many areas remained isolated with people in need of rescue. **[URL Here]**	Injured or dead people
@Username Tropical Storm Trami triggers killer floods, landslides in Philippines and the death toll is mounting READ MORE: **[URL Here]** **[URL Here]**	Injured or dead people

@Username Oh please pray to everyone who are affected of typhoon Kristine here in the Philippines 倮檹倮椢 hope everyone who needs help already rescue and no more hurt. The wind is so strong and heavy rain is here again 倮椢 please stop this already 倮檹倮橏 it's extreme alert in my place. **[URL Here]**	Sympathy and Support
@Username Dear God, we lift up the Bicol region to you during tropical storm #Kristine. The heavy rains have caused severe flooding, forcing thousands of families to evacuate. We pray for their safety and comfort. With several provinces under Signal #2, we ask for your divine protection鈥 **[URL Here]**	Sympathy and Support

Block 7 (Human)

Tweet	**Label**
@Username Parts of Taft Avenue in Manila flooded, but passable. [as of 2:15am Oct 25] #KristinePH #SelfieBalita **[URL Here]**	Caution and Advice
@Username (2/2) emphasizing that severe flooding across the provinces of Bicol, Batangas, Laguna, Rizal, and surrounding cities reflects the government鈥檚 failure to mitigate the impacts of typhoons through environmental protection.	Caution and Advice
@Username 2 typhoons battering the country right now and it's really unsafe to go. Manila is still under STS Signal no. 1 and many areas are flooded.	Caution and Advice
@Username @Username A landslide in Libon, Philippines, caused by Tropical Storm Trami, has resulted in severe damage. Rescue operations are ongoing.	Infrastructure and utility damage
@Username Not just Bicol region but even other provinces in North and South Luzon are currently experiencing heavy rainfall including Cavite due to Severe Tropical Storm. There are road closures already in most main roads of cavite due to floods and not passable to all type of vehicles. **[URL Here]**	Infrastructure and utility damage

Block 7 (LLM)

Tweet	Label
@Username Parts of Taft Avenue in Manila flooded, but passable. [as of 2:15am Oct 25] #KristinePH #SelfieBalita [**URL Here]**	Caution and Advice
@Username (2/2) emphasizing that severe flooding across the provinces of Bicol, Batangas, Laguna, Rizal, and surrounding cities reflects the government鈥檚 failure to mitigate the impacts of typhoons through environmental protection.	Caution and advice
@Username 2 typhoons battering the country right now and it's really unsafe to go. Manila is still under STS Signal no. 1 and many areas are flooded.	Caution and Advice
@Username @Username A landslide in Libon, Philippines, caused by Tropical Storm Trami, has resulted in severe damage. Rescue operations are ongoing.	Infrastructure and utility damage
@Username Not just Bicol region but even other provinces in North and South Luzon are currently experiencing heavy rainfall including Cavite due to Severe Tropical Storm. There are road closures already in most main roads of cavite due to floods and not passable to all type of vehicles.**[URL Here]**	Infrastructure and utility damage

Block 8 (Human)

Tweet	Label
@Username Don't usually do this but today I'm asking for prayers for my friends and family in the southern Philippines islands. The typhoon and now coastal tornado in my home town major damage and flooding . My family in the Deep South and in the Philippines affected by all the chaos	Sympathy and Support
@Username @Username Praying for those affected by the flooding. And for the families who have lost loved ones. 馃檹 馃檹 馃檹	Sympathy and Support

@Username PLEASE PRAY FOR BATANGAS, as some municipality are now flooding餘槴	Sympathy and Support
@Username In the Philippine, a man escaped being swept away by floodwater by clinging to a palm tree. Local residents shared footage on social media, seeking help from rescuers as heavy rains from tropical cyclone Trami (Kristine) hit the island of Luzon. The man was later rescued by his鈥**[URL Here]**	Rescue, volunteering, or donation effort
@Username LOOK: Former Vice President Leni Robredo braves waist-deep flood waters to deliver relief goods to those affected by severe tropical storm Kristine in Naga City. |餘撕 : Egine Oquindo Baral/Facebook via Alliah Jalimao, **[URL Here]**	Rescue, volunteering, or donation effort

Block 8 (LLM)

Tweet	**Label**
@Username Don't usually do this but today I'm asking for prayers for my friends and family in the southern Philippines islands. The typhoon and now coastal tornado in my home town major damage and flooding. My family in the Deep South and in the Philippines affected by all	Sympathy and Support
@Username @Username Praying for those affected by the flooding. And for the families who have lost loved ones. 餘檐 餘嚨餘嚒	Sympathy and Support
@Username PLEASE PRAY FOR BATANGAS, as some municipality are now flooding餘槴	Sympathy and Support

@Username In the Philippine, a man escaped being swept away by floodwater by clinging to a palm tree. Local residents shared footage on social media, seeking help from rescuers as heavy rains from tropical cyclone Trami (Kristine) hit the island of Luzon. The man was later rescued by his鈥 **[URL Here]**	Rescue, volunteering, or donation effort
@Username LOOK: Former Vice President Leni Robredo braves waist-deep flood waters to deliver relief goods to those affected by severe tropical storm Kristine in Naga City. |馃摳 : Egine Oquindo Baral/Facebook via Alliah Jalimao, #LeniRobredo #KristinePH **[URL Here]**	Rescue, volunteering, or donation effort

Block 9 (Human)

Tweet	**Label**
@Username @Username Heavy rain and flood in Philippines	Infrastructure and utility damage
@Username @Username the whole philippines is affected,many cancel flights in province,the road is flooding its not summer its rainy days and theres upcoming another typhoon maybe landing tonight or tommorrow,馃檹	Infrastructure and utility damage
@Username Many towns in Bicol, like Baao in Camarines Sur, remain flooded due to Severe Tropical Storm #KristinePH. Over 2 million people have been affected and lives have been lost. Disasters will persist if oil & gas companies keep damaging the environment. #MakeClimatePollutersPay **[URL Here]**	Infrastructure and utility damage
@Username At least 126 dead and missing in massive flooding, landslides in #Philippines **[URL Here]**	Injured or dead people

@Username The death toll and number of missing persons due to severe flooding and landslides caused by Tropical Storm Trami in the Philippines has nearly reached 130. The president announced on Saturday that many areas remain isolated, with people still requiring rescue. Trami moved away鈥 **[URL Here]**	Injured or dead people

Block 9 (LLM)

Tweet	**Label**
@Username @Username Heavy rain and flood in Philippines	Infrastructure and utility damage
@Username @Username the whole philippines is affected,many cancel flights in province,the road is flooding its not summer its rainy days and theres upcoming another typhoon maybe landing tonight or tommorrow,馃槶	Infrastructure and utility damage
@Username Many towns in Bicol, like Baao in Camarines Sur, remain flooded due to Severe Tropical Storm #KristinePH. Over 2 million people have been affected and lives have been lost. Disasters will persist if oil & gas companies keep damaging the environment. #MakeClimatePollutersPay **[URL Here]**	Infrastructure and utility damage
@Username At least 126 dead and missing in massive flooding, landslides in #Philippines	Injured or dead people
@Username The death toll and number of missing persons due to severe flooding and landslides caused by Tropical Storm Trami in the Philippines has nearly reached 130. The president announced on Saturday that many areas remain isolated, with people still requiring rescue.	Injured or dead people

Block 10 (Human)

Tweet	Label
@Username Residents of Lemery, Batangas were forced to evacuate due to severe flooding | Nagpapatrol, Jeff Caparas BNN_Breaking	Rescue, volunteering, or donation effort
@Username The local government of Pasig City said on Friday morning, October 25, that a total of 22 evacuation centers are open in the city, with 1,336 families temporarily sheltered due to the strong winds and heavy rains caused by Severe Tropical Storm #KristinePH.鈥	Rescue, volunteering, or donation effort
@Username #Philippine rescuers battle flood waters to reach stranded **[URL Here]**	Rescue, volunteering, or donation effort
@Username @Username As per his logic, floods in Philippines are caused by Indian rivers since some part of India gets flooded.	Caution and advice
@Username @Username They are not just affected by the flood, also there was a landslide happened in Antipolo and not safe to travel at this moment.	Caution and advice

Block 10 (LLM)

Tweet	Label
@Username Residents of Lemery, Batangas were forced to evacuate due to severe flooding | Nagpapatrol, Jeff Caparas	Rescue, volunteering, or donation effort
@Username The local government of Pasig City said on Friday morning, October 25, that a total of 22 evacuation centers are open in the city, with 1,336 families temporarily sheltered due to the strong winds and heavy rains caused by Severe Tropical Storm #KristinePH.鈥 **[URL Here]**	Rescue, volunteering, or donation effort

@Username #Philippine rescuers battle flood waters to reach stranded residents **[URL Here]**	Rescue, volunteering, or donation effort
@Username @Username As per his logic, floods in Philippines are caused by Indian rivers since some part of India gets flooded.	Caution and advice
@Username @Username They are not just affected by the flood, also there was a landslide happened in Antipolo and not safe to travel at this moment.	Caution and advice