Ghancaran
Jurnal Pendidikan Bahasa dan Sastra Indonesia
P-ISSN 2714-8955 | E-ISSN 2715-9132
https://ejournal.uinmadura.ac.id/index.php/ghancaran/index
https://doi.org/10.19105/ghancaran.v20i2.26173

OPEN ACCESS

RESEARCH ARTICLE

Reading disaster misinformation through the SAFE model: A critical digital literacy approach

Rahmad Nuthihar[1]*, Dadang S. Anshori[2], Isah Cahyani[3], Ixsir Eliya[2]

[1,2,3]Universitas Pendidikan Indonesia, Bandung, Indonesia
[4]Universitas Islam Negeri Fatmawati Sukarno Bengkulu, Bengkulu, Indonesia

***Corresponding Author**
r.nutihar@upi.edu

ABSTRACT
Information can spread quickly through digital media, especially social media, even when its accuracy has not been established. This creates room for misinformation, which may cause misunderstanding and public concern, particularly during disasters. This study explores the application of the SAFE Model, developed by Nuthihar et al. (2026), as a framework for analyzing misinformation. Originally developed for teaching popular article writing, the model is recontextualized into four analytical stages: Surveying, Arguing, Framing, and Evaluation. The data consist of six cases of disaster-related misinformation circulated in Indonesia in 2026 and documented on the TurnBackHoax website. The data were collected through documentation and analyzed qualitatively using the four stages of SAFE. The findings reveal several recurring patterns in the construction of disaster misinformation, including unsupported causal claims, miscontextualized visual evidence, unverified numerical information, and misleading references to institutional authority. For example, one post associated HAARP with natural disasters without scientific evidence, while another used an image of a 2025 fire in Brunei to support a claim about a 2026 forest fire in Kalimantan. Other cases presented unverified claims about the Lewotobi eruption, high waves in Gunungkidul, BMKG's dry-season prediction, and a mega-tsunami in Kuta, Bali. These findings demonstrate that disaster misinformation involves not only inaccurate claims but also the use of evidence and framing that may create an appearance of credibility. The study highlights the potential of SAFE as a critical digital literacy framework for examining how misinformation is constructed, presented, and verified.

Keywords: critical digital literacy; disaster misinformation; discourse analysis; SAFE Model; social media

ARTICLE HISTORY
Received: September 16, 2026
Accepted: October 28, 2026
Published: November 14, 2026

LICENSE
Copyright © 2026 The Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) License.
https://creativecommons.org/licenses/by/4.0/.

How to Cite This Article
Nuthihar, R., et al. (2026). Reading disaster misinformation through the SAFE model: A critical digital literacy approach. *Ghancaran: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, Special Edition Lalongét VII, 678-697.
https://doi.org/ghancaran/v20i2.26173

INTRODUCTION

The use of digital and social media has transformed how information is produced, accessed, and distributed. Information can circulate widely within a short time, even without sufficient verification, allowing misinformation and disinformation to spread alongside factual content (Søe, 2021). Such information may appear in textual, visual, or audiovisual forms, which can make misleading messages seem convincing (Rubin, 2019). Social media users also contribute to this circulation through sharing and commenting, extending the reach of information beyond its original source (Savira et al., 2021). Consequently, evaluating digital information requires attention not only to factual accuracy but also to how messages are constructed and presented to establish credibility (Luo et al., 2022).

This issue becomes particularly important when misinformation concerns disasters. Misleading information may involve the location, timing, magnitude, or impact of earthquakes, tsunamis, volcanic eruptions, forest fires, and extreme weather events (Hilberts et al., 2025; T & Mathew, 2022). Images, videos, numerical details, and references to official institutions may make such claims appear credible, even when their context or accuracy has not been established. In disaster situations, misleading information can create public confusion and influence decisions concerning potential threats. Examining disaster misinformation therefore requires attention to the claims themselves, the evidence presented, and the linguistic and visual features that shape their interpretation (Molina et al., 2021).

Critical digital literacy is relevant to this challenge because readers need to identify, analyze, evaluate, and verify information encountered online (Eliya et al., 2026; Knight et al., 2023; Zubiaga et al., 2016). In this study, critical digital literacy refers to the ability to examine digital information by considering its content, sources, argumentative structures, linguistic and visual representations, and contextual accuracy. This understanding draws on Valle et al. (2025), who emphasize users' goals, contexts, inquiry, reflection, and action in social media literacy, who highlight information acquisition, critical evaluation, and evidence-based reasoning. Accordingly, four analytical indicators guide this study: (1) message identification and contextual awareness; (2) argumentative evaluation; (3) critical framing analysis; and (4) information verification. These indicators provide the conceptual basis for adapting SAFE into Surveying, Arguing, Framing, and Evaluation.

Argumentation offers a way to examine how digital claims are supported. The Claim–Evidence–Reasoning (CER) framework identifies the claim, the evidence presented in its support, and the reasoning connecting them (Talitha et al., 2026), social media posts, this

reasoning may be implicit rather than explicitly stated. Images, videos, numbers, and institutional references can function as supporting material when presented alongside a claim. Their argumentative role can be examined through captions, text overlays, visual sequences, and narrative context. However, when the available content does not establish a connection between the claim and the accompanying material, the reasoning should be classified as indeterminate rather than inferred. This distinction is important because the presence of evidence does not necessarily mean that it adequately supports a claim (Pimentel & Campbell, 2025).

The analysis also draws on *The 6+1 Traits of Writing* (Marangoz & Ensar, 2025), which comprises Ideas, Organization, Voice, Word Choice, Sentence Fluency, Conventions, and Presentation. Originally developed to guide writing production and assessment, these dimensions are recontextualized in this study to examine how digital messages are constructed. Ideas and Organization help identify the central message and arrangement of information, while Voice, Word Choice, and Sentence Fluency direct attention to tone and language. Conventions and Presentation extend the analysis to the final form of verbal and visual materials. The framework is not used to determine factual accuracy but to examine how message construction shapes meaning and persuasive force.

These two foundations inform the SAFE Model developed by Nuthihar et al. (2026) for teaching students to write popular articles. Its original stages are Surveying, Arguing, Framing-Writing, and Editing. The model integrates *The 6+1 Traits of Writing* as guidance for text quality and CER as a basis for constructing arguments. Surveying involves identifying issues and sources; Arguing develops claims, evidence, and reasoning; Framing-Writing organizes ideas into a coherent text; and Editing reviews and improves the draft. Thus, the original SAFE Model is oriented toward writing production rather than the analysis of existing digital messages.

Although fact-checking establishes whether claims correspond with available facts, it does not necessarily explain how misleading messages acquire credibility. Such messages may combine questionable evidence, weak reasoning, particular language choices, and visual materials to influence readers' interpretations (Volkova & Jang, 2018). This creates a need for an analytical framework that connects message identification, argument analysis, framing, and verification within a single reading process.

This study therefore recontextualizes SAFE from a writing instruction model into an analytical framework for critical digital literacy. The adapted model retains the four-stage structure but changes Framing-Writing to Framing and replaces Editing with Evaluation. The

latter examines claims, evidence, sources, and contextual information against fact-checking records and other verifiable sources. Surveying identifies what a message says and shows; Arguing examines its claims, evidence, and reasoning; Framing analyzes how the message is presented; and Evaluation checks its factual basis. The adaptation shifts SAFE from producing and revising texts to critically examining digital information.

The study applies this adapted model to six cases of disaster misinformation circulated through social media in Indonesia. It examines how claims, evidence, reasoning, and framing are constructed, as well as what their analysis reveals about SAFE's potential as a critical digital literacy framework. Specifically, the study addresses two research questions: (RQ1) How are claims, evidence, reasoning, and framing constructed in disaster misinformation through the SAFE Model? and (RQ2) What does the application of the SAFE Model reveal about its potential as a critical digital literacy framework?

METHOD

This study employed a qualitative multiple-case study involving six cases of disaster-related misinformation circulated on social media in Indonesia during 2026. Each case was examined individually because the claims, supporting materials, narratives, visual content, and factual contexts differed. The analysis focused on how misinformation was constructed rather than how frequently it occurred. The SAFE Model was used as an analytical framework, not as an intervention whose effectiveness was statistically tested.

The data comprised six social media posts from Instagram, Facebook, TikTok, and YouTube. These posts constituted the primary data and included written narratives, captions, images, and videos. Secondary data consisted of fact-checking reports published on TurnBackHoax, managed by the Indonesian Anti-Slander Society (MAFINDO), and other verifiable sources used to establish the factual context of each case. Table 1 presents the six cases and their corresponding data sources.

Case	Disaster issue	Platform	Data form	Fact-checking source
D1	Volcanic eruption and HAARP	Instagram	Video and text	Mafindo (2026b)
D2	Forest fire allegedly spreading from Kalimantan to Brunei	Facebook	Image and text	Mafindo (2026d)
D3	Lewotobi Laki-Laki eruption	TikTok	Video and text	Mafindo (2026c)
D4	Tsunami/high waves in Gunungkidul	YouTube	Video and text	Mafindo (2026e)
D5	Claim of Indonesia's worst dry season in 30 years	Facebook	Image and text	Mafindo (2026a)
D6	Alleged mega-tsunami in Kuta, Bali	Facebook	Video and text	Mafindo (2026f)

Table 1. Dataset of Disaster Misinformation Cases (Source: https://turnbackhoax.id/)

The cases were selected purposively based on four criteria: (1) the misinformation concerned a disaster or disaster risk in Indonesia; (2) the post circulated on social media during 2026; (3) it contained an identifiable claim and analyzable textual or visual material; and (4) a corresponding fact-checking report was available on TurnBackHoax. For each case, the original post or an archived copy was documented, including its account, publication date, narrative, images, and videos. Engagement figures were recorded when available but were used only as contextual information, not as evidence of credibility or audience influence.

Each individual social media post constituted one unit of analysis, resulting in six cases. Its textual and visual components, including claims, captions, narratives, images, videos, numerical details, and institutional references, were treated as analytical elements rather than separate units. TurnBackHoax reports and other verifiable sources served as supporting materials for contextualizing and checking the claims.

The six cases were analyzed using the adapted SAFE Model, comprising *Surveying, Arguing, Framing,* and *Evaluation*. In *Surveying*, the analysis identified the main message, context, and textual or visual materials in each post. *Arguing* examined the claim, the evidence presented, and the reasoning connecting them. When reasoning was implicit, it was interpreted from the relationships established through captions, text overlays, visual sequences, and narrative context. If the available material was insufficient to establish such a relationship, the reasoning was classified as indeterminate. *Framing* examined how language and visual elements represented the event, including tone, word choice, sentence construction, organization, and presentation. Finally, *Evaluation* checked the claims and supporting materials against fact-checking reports and other verifiable sources, considering their sources, context, dates, and visual authenticity where relevant.

The analytical dimensions were adapted from the original SAFE Model. *The 6+1 Traits of Writing* informed the analysis of message construction across *Surveying, Arguing,* and *Framing* (Nuthihar et al., 2026). Meanwhile, Claim, Evidence, and Reasoning (CER) guided the examination of argumentative relationships in the *Arguing* stage (Clark & Sampson, 2008; Sampson & Clark, 2008a). *Evaluation* relied on external verification rather than the 6+1 Traits. The operationalization of these dimensions is presented in Table 2.

SAFE stage	Analytical dimensions	Analytical focus
Surveying	Ideas, Presentation	Message, context, and mode of presentation
Arguing	Ideas, Organization, and CER	Claim, evidence, reasoning, and argumentative relationships

Framing	Voice, Word Choice, Sentence Fluency, Organization, Presentation	Linguistic and visual construction of credibility, urgency, and interpretation
Evaluation	External verification	Verification of claims, evidence, sources, context, time, and visual authenticity

Table 2. SAFE Analytical Framework for Disaster Misinformation

Data analysis involved two stages. First, each case was examined individually through the four SAFE stages to preserve its specific context. Second, a cross-case comparison identified recurring features and differences in how claims were presented, supported, and framed, as well as how the verification results confirmed or contradicted the information in the posts. The findings were limited to the six selected cases and were not intended to represent disaster misinformation in general.

RESULTS AND DISCUSSION

The cases examined here involve different disaster claims in Indonesia. They include a possible link between HAARP and natural disasters, a forest fire said to have reached Brunei from Kalimantan, the Lewotobi eruption, high waves in Gunungkidul, a warning about the 2026 dry season, and a reported mega-tsunami in Bali. These claims appeared in different forms on social media. Some were stated as reports of events that had already occurred. Others relied on speculation or referred to an official institution to give weight to the information being shared. Table 3 presents the main details of each case, including the platform, date, claim, and form of presentation. Likes, comments, and shares are also reported when available. These figures are used only to provide context for the circulation of the posts, not to assess their credibility or audience response.

Case	Platform	Disaster issue	Main claim
D1	Instagram	Volcanic activity and HAARP	A series of natural disasters was questioned as being entirely natural and linked to the possible involvement of HAARP.
D2	Facebook	Forest fire	A forest fire in Kalimantan was claimed to have reached Brunei Darussalam and burned residents' homes.
D3	TikTok	Lewotobi Laki-Laki eruption	The eruption of Mount Lewotobi Laki-Laki was reported to have reached 10 km while NTT was still recovering from an earthquake.
D4	YouTube	Tsunami/high waves	Waves reaching 10 meters were described as the "Gunung Kidul Tsunami" and claimed to have occurred along the southern coast and in five other cities.
D5	Facebook	Dry season	BMKG was claimed to have warned that the 2026 dry season would be the most severe in the past 30 years.
D6	Facebook	*Mega-tsunami*	A *mega-tsunami* was claimed to have destroyed Kuta, Bali, including roads and resorts.

Table 3. Overview of Disaster Misinformation Cases (Source: Authors' own elaboration)

The six cases differ in both the type of disaster and the way the information was presented. Most of the posts combined verbal claims with images or videos. The HAARP case was somewhat different because it relied on a longer personal narrative. Numbers, locations, institutional names, and particular expressions also appeared in several posts to describe the scale or circumstances of the disaster. Figure 1 presents the original form of the six posts so that the narrative and visual elements examined in the analysis can be seen together.



Figure 1. The six disaster misinformation cases examined in the study: (a) HAARP and volcanic activity; (b) the Kalimantan forest fire and Brunei Darussalam; (c) the Lewotobi Laki-Laki eruption; (d) tsunami/high waves in Gunungkidul; (e) BMKG and the 2026 dry season; and (f) the alleged mega-tsunami in Kuta, Bali (Source: https://turnbackhoax.id/)

Surveying Disaster Misinformation: Identifying Claims, Sources, and Message Characteristics

The *Surveying* stage identified the claims and presentation characteristics of the six social media posts using five indicators: claim, source, platform, context, and message characteristics. These indicators covered the main statement, the post's origin and platform, the disaster event and circumstances referred to, and the textual, visual, numerical, and institutional elements used to communicate the claim. At this stage, the analysis documented the available information without assessing claim accuracy or visual authenticity. Table 4 summarizes the identification results.

Case	Claim	Source and Platform	Context	Message Characteristics
D1	HAARP is presented as a possible explanation for natural disasters, including volcanic activity.	Original post on Instagram; (Mafindo, 2026b).	HAARP is introduced as an alternative explanation for natural phenomena in Indonesia.	Video and personal narrative referring to conspiracy theories, the US Air Force, US Navy, and DARPA.
D2	A forest fire from Kalimantan has spread to Brunei Darussalam and burned residents' homes.	Original post on Facebook; Mafindo (2026d).	An alleged cross-border fire affecting residential areas in Brunei.	Fire image accompanied by text identifying Kalimantan as the origin and Brunei as the affected location.
D3	Mount Lewotobi Laki-Laki erupted with an ash column reaching 10,000 meters while NTT was recovering from an earthquake.	Original post on TikTok; Mafindo (2026c).	The alleged eruption is connected to the post-earthquake situation in East Nusa Tenggara (NTT).	Eruption video, "10KM/10,000 meters," "VIRAL," and "INDONESIA TIMUR KEMBALI BERKABUNG."
D4	Waves reaching 10 meters in Gunungkidul are described as a tsunami.	Original post on YouTube; Mafindo (2026e).	An alleged recent coastal disaster in Gunungkidul.	High-wave video with "Tsunami Gunung Kidul," "PANTAI SELATAN NAIK," "Baru Saja," and "10 Meter."
D5	BMKG warned that Indonesia's 2026 dry season would be the most severe in 30 years.	Original post on Facebook; Mafindo (2026a).	A national seasonal forecast attributed to BMKG.	Image and text featuring BMKG's name, a dry-season prediction, and "30 years."
D6	A mega-tsunami has destroyed Kuta, Bali.	Original post on Facebook; Mafindo (2026f).	An alleged mega-tsunami and its destructive impact on buildings in Kuta.	Video of large waves striking buildings, accompanied by a textual claim identifying the event and location.

Table 4. Identification of the Six Disaster Misinformation Cases through the Surveying Stage

The six cases presented different disaster-related claims. D1 associated HAARP with natural disasters; D2 described a forest fire allegedly spreading from Kalimantan to Brunei; D3 connected the Lewotobi eruption with a 10,000-meter ash column; D4 described 10-

meter waves in Gunungkidul as a tsunami; D5 attributed a severe dry-season prediction to BMKG; and D6 claimed that a mega-tsunami had destroyed Kuta, Bali. These claims appeared across four platforms: Instagram (D1), Facebook (D2, D5, and D6), TikTok (D3), and YouTube (D4). Their corresponding fact-checking reports were published by Mafindo through TurnBackHoax. This distribution describes the selected cases and does not indicate the prevalence of misinformation on each platform.

The posts also situated their claims in different geographical and institutional contexts. These included HAARP and volcanic activity in D1, an alleged cross-border fire in D2, the Lewotobi eruption and the earthquake aftermath in NTT in D3, the southern coast of Gunungkidul in D4, a national seasonal forecast attributed to BMKG in D5, and Kuta, Bali, in D6. Such references connected the claims to identifiable locations, institutions, and disaster circumstances.

In terms of message characteristics, the posts combined textual and visual elements in different ways. D1 used personal narrative and references to international institutions. D2 paired a fire image with a cross-border claim. D3 used eruption footage, numerical information, and emotionally charged expressions. D4 combined high-wave footage with references to a tsunami, recency, and wave height. D5 used BMKG's name and a 30-year comparison, while D6 presented large-wave footage as evidence of a mega-tsunami in Bali.

These findings show how images, videos, numbers, institutional names, and textual expressions were used to communicate disaster claims. However, their presence does not establish the accuracy of the claims or the validity of the supporting materials. Their argumentative function is examined in the *Arguing* stage, their linguistic and visual construction in *Framing*, and their accuracy and authenticity through external verification in *Evaluation*.

Arguing: Claims, Evidence, and Reasoning

The *Arguing* stage examined the *claim*, the evidence presented to support it, and the reasoning connecting them. A claim refers to the statement readers are expected to accept, evidence refers to the material presented in its support, and reasoning concerns the relationship established between the two. Across the six posts, the supporting materials included textual statements, images, videos, numerical information, and institutional references. Table 5 summarizes these components.

Case	Claim	Evidence presented	Reasoning

D1	HAARP is presented as a possible explanation for natural disasters.	Personal narrative, references to HAARP, the US Air Force, US Navy, and DARPA, and assertions about HAARP's alleged capabilities.	Institutional associations and alleged capabilities suggest a possible causal connection between HAARP and natural disasters.
D2	The Kalimantan forest fire has reached Brunei and burned residents' homes.	A fire image and the statement *"kebakaran hutan Kalimantan mulai masuk negara Brunei"* [the Kalimantan forest fire has begun to enter Brunei].	The image is assumed to represent the fire and its alleged cross-border impact, although its location and origin are not demonstrated.
D3	The Lewotobi eruption produced an ash column reaching 10,000 meters.	Eruption footage and the numerical expression "10KM/10,000 meters."	The video represents the alleged eruption, while the figure indicates its claimed magnitude.
D4	Gunungkidul experienced a tsunami or tsunami-like waves reaching 10 meters.	High-wave footage and the expressions *"Tsunami Gunung Kidul," "PANTAI SELATAN NAIK," "Baru Saja"* [Just Now], and "10 Meter."	The text assigns a location, time, event type, and height to the footage.
D5	BMKG warned that the 2026 dry season would be the most severe in 30 years.	BMKG's name and the statement *"kemarau tahun 2026 di Indonesia bakal jadi kemarau terparah dalam 30 tahun terakhir"* [the 2026 dry season in Indonesia will be the most severe in the past 30 years].	Institutional attribution and historical comparison lend apparent authority to the prediction.
D6	A mega-tsunami has destroyed Kuta, Bali.	Footage of large waves striking buildings and the statement *"Mega tsunami hancurkan Kuta, Bali"* [Mega-tsunami destroys Kuta, Bali].	The video is presented as a representation of the alleged disaster, while the text identifies its location and impact.

Table 5. Claim, Evidence, and Reasoning Structures in Disaster Misinformation

D1 develops its claim through the uploader's personal narrative, beginning with *"gak mau terima dengan penjelasan yang biasa dan lumrah HARUS diterima"* [do not want to accept the usual and commonly accepted explanation]. HAARP is introduced as an *"alternative lain"* [another alternative], alongside references to the US Air Force, US Navy,

and DARPA. The post also acknowledges *"Walau tdk ada bukti ilmiah"* [although there is no scientific evidence], before associating HAARP with the ability to *"mengendalikan cuaca, membuat badai hingga gempa bumi"* [control the weather, create storms, and even cause earthquakes]. Its reasoning moves from institutional associations to a possible causal connection with natural disasters. However, the post provides no scientific studies or technical documentation demonstrating this connection.

D2, D3, D4, and D6 rely primarily on visual materials accompanied by textual or numerical information. D2 presents a fire image alongside the claim that a Kalimantan forest fire has entered Brunei and burned residents' homes. The implied reasoning assumes that the image depicts the event and location described, although the image itself does not establish its origin or geographical context. D3 uses eruption footage and the figure "10KM/10,000 meters" to represent the alleged Lewotobi eruption and its magnitude. However, the post provides no information establishing the footage's origin or how the height was determined. Similarly, D4 combines high-wave footage with *"Tsunami Gunung Kidul," "PANTAI SELATAN NAIK," "Baru Saja"* [Just Now], and "10 Meter." These expressions assign a location, time, event type, and wave height to the video, although the footage alone cannot establish these details. D6 presents large waves striking buildings alongside *"Mega tsunami hancurkan Kuta, Bali"* [Mega-tsunami destroys Kuta, Bali]. The text identifies the footage as a destructive event in Kuta, but the video alone does not establish its location, timing, or event type. D5 uses institutional attribution and numerical comparison. The post attributes the prediction of the most severe dry season in 30 years to BMKG. This attribution establishes apparent institutional authority, while the 30-year comparison gives the claim a historical basis. However, neither the original BMKG statement nor supporting historical data is provided in the post.

Across the six cases, evidence took different forms, including personal narratives, institutional references, images, videos, and numerical information. Although these materials were presented to support the claims, their presence does not necessarily establish their validity or relevance. The *Arguing* stage therefore examined both the materials presented and the explicit or implicit reasoning connecting them to the claims. Their accuracy, authenticity, and contextual relevance were subsequently examined in the *Evaluation* stage, while their linguistic and visual construction was analyzed in *Framing*.

Framing: Making Disaster Misinformation Believable

The *Framing* stage examined how linguistic and visual elements were used to make disaster-related claims appear convincing. The analysis focused on four features: sensationalist language, numerical details without clear supporting sources, references to official institutions, and images or videos presented as representations of specific disaster events. Table 6 summarizes the framing features identified in the six cases.

Case	Voice and Word Choice	Sentence Fluency and Organization	Presentation	Framing Pattern
D1	*"saya orangnya skeptis banget"* [I am very skeptical]; *"teori konspirasi"* [conspiracy theory]; "fakta yang tertunda" [delayed fact]	Moves from questioning conventional explanations to presenting HAARP as an alternative.	References to the US Air Force, US Navy, and DARPA.	Skepticism and authority
D2	*"mulai masuk negara Brunei"* [has begun to enter Brunei]; *"bakar rumah penduduk"* [burned residents' homes]	Presents a direct relationship between the alleged fire and its impact.	Fire image accompanying the cross-border claim.	Threat and proximity
D3	*"ditimpa bencana lagi"* [struck by another disaster]; "VIRAL"; *"KEMBALI BERKABUNG"* [MOURNING AGAIN]	Presents the eruption as a continuation of previous disasters.	Eruption footage, capital letters, and "10KM/10,000 meters."	Threat and repeated disaster
D4	"Tsunami Gunung Kidul"; *"Baru Saja"* [Just Now]; "10 Meter"	Uses short expressions to emphasize event type, timing, and scale.	High-wave video, capital letters, and numerical information.	Urgency and threat
D5	*"BMKG ingatkan"* [BMKG warns]; *"terparah dalam 30 tahun*	Places the institutional attribution before the prediction.	BMKG's name and "30 years" are emphasized.	Authority and severity

	terakhir" [most severe in the past 30 years]			
D6	"mega tsunami"; *"hancurkan Kuta, Bali"* [destroys Kuta, Bali]	Uses a short, direct statement emphasizing destruction.	Video of large waves striking buildings.	Catastrophe and visual proximity

Table 6. Framing Elements in Disaster Misinformation (Source: Authors' own elaboration)

Sensationalist language is particularly evident in D3, D4, and D6. D3 uses *"ditimpa bencana lagi"* [struck by another disaster] and *"INDONESIA TIMUR KEMBALI BERKABUNG"* [EASTERN INDONESIA MOURNS AGAIN] to frame the eruption as part of continuing disasters. D4 combines "Tsunami Gunung Kidul," *"Baru Saja"* [Just Now], and "10 Meter" to suggest an urgent and large-scale event. Similarly, D6 uses *"mega tsunami hancurkan Kuta, Bali"* [mega-tsunami destroys Kuta, Bali] to emphasize catastrophic impact. These expressions frame the claims through urgency, threat, and destruction.

Numerical details also create an appearance of precision. D3 presents "10KM/10,000 meters" for the alleged ash column, D4 uses "10 Meter" for the waves, and D5 refers to "30 years" in its dry-season prediction. However, the posts do not provide identifiable sources or supporting records for these figures. Their apparent specificity therefore cannot be taken as evidence of accuracy.

Institutional references function as authority cues in D1 and D5. D1 begins with the personal statement *"saya orangnya skeptis banget"* [I am very skeptical] and questions conventional explanations before introducing HAARP as an alternative. References to the US Air Force, US Navy, and DARPA appear within this narrative, potentially lending institutional credibility to the proposed explanation. D5 more directly attributes its prediction to BMKG through *"BMKG ingatkan"* [BMKG warns], followed by the claim that the dry season will be the most severe in 30 years. Neither post, however, provides verifiable evidence that the named institutions support the claims.

Visual materials further reinforce the messages in D2, D3, D4, and D6. D2 pairs a fire image with a claim about a fire spreading from Kalimantan to Brunei. D3 combines eruption footage with a numerical figure, while D4 labels high-wave footage as a recent tsunami in Gunungkidul. D6 presents large waves striking buildings as a mega-tsunami in Kuta, Bali. These combinations encourage readers to associate the visuals with the specific events described in the accompanying text.

The *Evaluation* findings reveal discrepancies between these visuals and their claimed contexts. The D2 image originated from a fire in Brunei in 2025, not the alleged 2026 Kalimantan forest fire (Mafindo, 2026d). D3 used footage of an earlier Lewotobi eruption on 17 June 2025 (Mafindo, 2026c). D4 combined videos from different locations and times rather than documenting a single tsunami event in Gunungkidul (Mafindo, 2026e). The D6 video was identified in the fact-checking report as likely AI-generated (Mafindo, 2026f).

Overall, the six cases show how sensationalist language, numerical details, institutional references, and visual materials shape the presentation of disaster claims. These features can create impressions of urgency, precision, authority, and visual authenticity without establishing factual accuracy. The *Framing* stage explains how the messages are made to appear believable, while *Evaluation* examines whether their claims and supporting materials are consistent with verifiable information.

Evaluation: From Claims to Factual Verification

The *Evaluation* stage examined the accuracy of claims and the validity of their supporting materials by comparing them with fact-checking reports and other verifiable sources. The verification covered the claims, images, videos, numerical details, locations, dates, and institutional attributions identified in the previous stages. Table 7 summarizes the results.

Case	Claim examined	Verification results	Assessment
D1	HAARP may cause natural disasters, including earthquakes and volcanic activity.	HAARP is an ionospheric research program. No scientific evidence supports the claim that it can trigger earthquakes or volcanic eruptions.	Unsupported by scientific evidence (Mafindo, 2026b).
D2	A forest fire from Kalimantan reached Brunei and burned residents' homes.	The image originated from a fire in Brunei in 2025, not the 2026 Kalimantan forest fire. The reported cross-border impact involved haze, not fire reaching Brunei.	False and based on an image taken out of context (Mafindo, 2026d).
D3	The Lewotobi eruption produced a 10,000-meter ash column in the context described.	The video showed an eruption on 17 June 2025, not the event described in the post. The 10,000-meter figure could not be verified	Misleading event context; the numerical claim remains unverified (Mafindo, 2026c).

		from the available fact-checking evidence.	
D4	A tsunami or 10-meter waves occurred in Gunungkidul.	BPBD DIY confirmed no tsunami at Pantai Kukup during the relevant period. The video combined footage from different locations and times. BMKG forecast waves of 2.5–4 meters for 1–4 July 2026, with no tsunami warning.	False (Mafindo, 2026e).
D5	BMKG warned that the 2026 dry season would be the most severe in 30 years.	BMKG denied issuing the statement. Its forecast indicated below-normal rainfall, while 1997, 2005, 2015, and 2019 were drier than the 2026 forecast.	False attribution and inaccurate prediction (Mafindo, 2026a).
D6	A mega-tsunami destroyed Kuta, Bali.	No credible information confirmed the event. The video contained visual anomalies and was classified by Hive Moderation as likely AI-generated, with a 99.9% probability score.	Unsubstantiated event claim; video likely AI-generated (Mafindo, 2026f).

Table 7. Evaluation Results for Disaster Misinformation Claims

The findings reveal different verification problems across the six cases. D1 involves an unsupported causal connection: the existence of HAARP and its institutional associations do not demonstrate that it can cause natural disasters. D2 involves visual material taken out of context, as the image originated from a different fire in Brunei in 2025. These cases show that verification must examine both the claim and the material used to support it.

D3 and D4 demonstrate the importance of checking the origin, timing, and geographical context of visual materials. The D3 footage came from an earlier Lewotobi eruption, while the stated ash-column height remained unverified. In D4, footage from different locations and times was combined to represent a tsunami in Gunungkidul. The available official information did not support that representation. D5 and D6 present different verification issues. In D5, BMKG denied issuing the attributed warning, and its actual forecast did not support the claim of the most severe dry season in 30 years. In D6, no credible evidence confirmed the alleged mega-tsunami. Although the detection tool indicated a 99.9% probability of AI generation, this automated result should not be treated as conclusive forensic proof.

Overall, *Evaluation* identifies not only whether a claim is supported but also where its supporting materials become problematic. The discrepancies may concern causal explanations, visual origins, numerical details, institutional attributions, or event contexts. By checking these elements against verifiable information, SAFE encourages readers to move beyond accepting a message at face value and examine precisely which parts of it are inaccurate, unsupported, or unverifiable.

The SAFE Model as a Critical Digital Literacy Framework

The analysis demonstrates how the SAFE Model can be applied to examine disaster misinformation through four connected stages. *Surveying* identifies the messages and contexts presented in a post. *Arguing* examines the relationships among claims, evidence, and reasoning. *Framing* analyzes how linguistic, numerical, institutional, and visual elements shape the presentation of a message. Finally, *Evaluation* verifies claims and supporting materials against fact-checking reports and other reliable sources. Figure 2 illustrates these stages and their analytical focus.

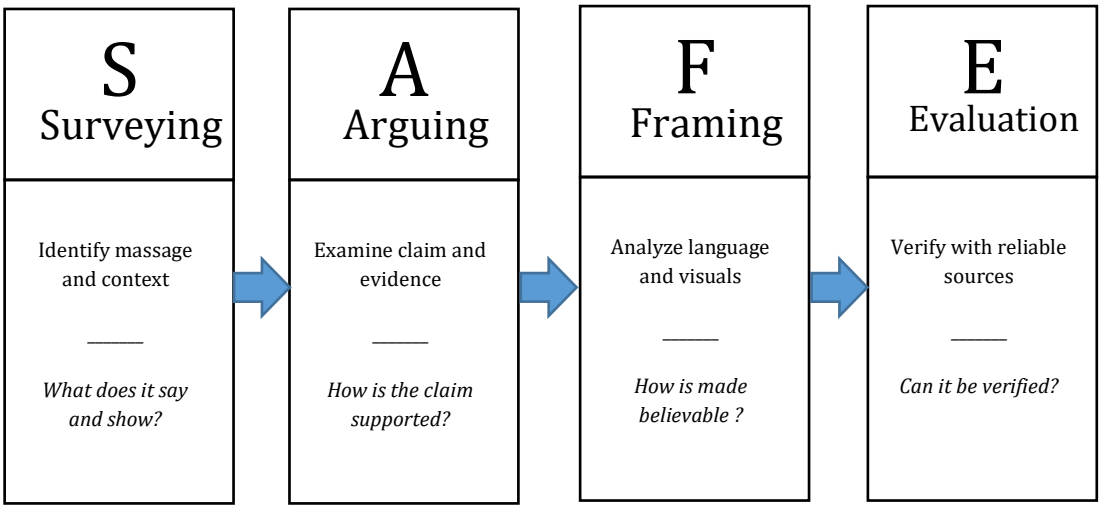


Figure 2. The SAFE Model as a Critical Digital Literacy Framework (*Source: Authors' own elaboration*)

The six cases show that misinformation may involve more than inaccurate claims. Problems can also occur in the sources, numerical details, and visual materials used to support them. In D2, the fire image originated from an unrelated incident in Brunei in 2025. D3 used footage from an earlier Lewotobi eruption, while D4 combined videos from different locations and times. D5 attributed a dry-season prediction to BMKG that the institution had denied issuing. D1 presented a causal connection between HAARP and natural disasters without scientific support, while D6 contained video material identified as likely AI-

generated (Mafindo, 2026a, 2026b, 2026c, 2026d, 2026e, 2026f). These findings illustrate the need to connect message identification, argumentative analysis, framing, and factual verification.

The four stages also demonstrate that verification is not merely a process of labeling information as true or false. Findings from *Evaluation* can clarify how materials identified in earlier stages were used to construct a claim. For example, tracing the D2 image to an unrelated fire reveals how visual evidence can be detached from its original context. Similarly, identifying the origins of the D3 footage and the combination of videos in D4 highlights the importance of examining visual materials independently from their accompanying claims.

SAFE's conceptual contribution to critical digital literacy lies in integrating four analytical questions: What does the message say and show? How is the claim supported? How is it made believable? Can it be verified? These questions encourage readers to examine not only the accuracy of information but also how claims are constructed and presented through textual, visual, numerical, and institutional elements. The question *"How was I made to believe it?"* therefore complements *"Is this information true?"* by directing attention to the construction of credibility.

This study presents SAFE as a conceptual and analytical framework for examining disaster misinformation. However, its effectiveness in improving readers' critical digital literacy has not been empirically tested. Further research is needed to investigate its pedagogical application and learning outcomes.

CONCLUSION

The analysis of the six cases reveals recurring patterns in the construction of disaster misinformation on social media. These include speculative causal claims, visual materials presented without sufficient contextual information, numerical details that lack verifiable support, and the use of institutional names to establish apparent credibility. The *Evaluation* stage further showed that some materials initially presented as evidence were unrelated to the events claimed, while certain institutional attributions and numerical statements could not be confirmed through the available sources. Taken together, these findings demonstrate that disaster misinformation does not rely solely on inaccurate claims but can also draw credibility from the way textual, visual, numerical, and institutional elements are combined.

The conceptual contribution of this study lies in recontextualizing SAFE from a writing instruction model into an analytical framework for critical digital literacy. Its four stages, *Surveying, Arguing, Framing,* and *Evaluation*, provide an integrated approach to examining the content of digital messages, the claims and evidence they present, the ways they are framed, and their factual validity. Rather than focusing exclusively on whether information is true or false, SAFE also draws attention to how a message is constructed to appear credible. The question *"How was I made to believe it?"* captures this critical orientation and highlights the importance of examining the relationship between claims, evidence, reasoning, and the linguistic and visual features of digital information.

This study is limited by its purposive selection of six cases and its reliance on available fact-checking records and other verifiable sources. It does not examine audience reception or empirically test SAFE as a digital literacy tool. Therefore, its findings provide an exploratory account of disaster misinformation and establish a conceptual basis for applying SAFE to its analysis, rather than demonstrating the model's effectiveness in improving readers' digital literacy. Future research could examine a broader range of misinformation across platforms and empirically investigate how readers use SAFE to identify, analyze, and verify digital information.

ACKNOWLEDGMENTS

The authors thank *Lembaga Pengelola Dana Pendidikan* (LPDP), The Finance Ministry of Indonesia, for supporting the publication of this article.

DECLARATIONS

Ethical considerations

Ethical approval was obtained for the earlier study in which the SAFE Model was developed with student participants (No. 260/UN40.K/PT.01.01/2026, Research Ethics Committee, Universitas Pendidikan Indonesia). This article uses a different type of data. It examines misinformation cases documented online and does not collect data from people or animals. For this reason, no additional ethical approval was sought for the textual analysis reported here.

Use of artificial intelligence (AI)

The authors used ChatGPT Plus (OpenAI) to assist with English grammar and paraphrasing during manuscript preparation. All AI-assisted text was reviewed and revised by the authors, who take full responsibility for the final content of the manuscript.

Conflict of Interest

No conflicts of interest.

Funding

The publication of this article was supported by the *Lembaga Pengelola Dana Pendidikan* (LPDP), Ministry of Finance, Republic of Indonesia.

REFERENCE

Eliya, I., Lida, U. M., Nuthihar, R., & Nurpadillah, V. (2026). Bahasa sebagai Ruang Penyembuhan: Representasi Wacana Kesehatan Mental bagi Muslimah di Media Sosial Instagram. *GHANCARAN: Jurnal Pendidikan Bahasa Dan Sastra Indonesia2*, *7*(2), 465–484. https://doi.org/10.19105/ghancaran.v7i2.22430

Hilberts, S., Govers, M., Petelos, E., & Evers, S. (2025). The Impact of Misinformation on Social Media in The Context of Natural Disasters: A Narrative Review. *JMIR Infodemiology*, *5*. https://doi.org/10.2196/70413

Knight, J., Dooly, M., & Barberà, E. (2023). Getting Smart: Towards Critical Digital Literacy Pedagogies. *Social Semiotics*, *33*(2), 326–349. https://doi.org/10.1080/10350330.2020.1836815

Luo, M., Hancock, J. T., & Markowitz, D. M. (2022). Credibility Perceptions and Detection Accuracy of Fake News Headlines on Social Media: Effects of Truth-Bias and Endorsement Cues. *Communication Research*, *49*(2), 171–195. https://doi.org/10.1177/0093650220921321

Mafindo. (2026a). *[Salah] BMKG: Kemarau 2026 Bakal Jadi yang Terparah dalam 30 Tahun Terakhir*. TurnBackHoax. https://turnbackhoax.id/articles/33656-salah-bmkg-kemarau-2026-bakal-jadi-yang-terparah-dalam-30-tahun-terakhir

Mafindo. (2026b). *[Salah] Erupsi Gunung Di Indonesia Berkaitan dengan HAARP*. TurnBackHoax. https://turnbackhoax.id/articles/36590-salah-erupsi-gunung-di-indonesia-berkaitan-dengan-haarp

Mafindo. (2026c). *[Salah] Gunung Lewotobi Laki-Laki Erupsi Usai Gempa NTT*. TurnBackHoax. https://turnbackhoax.id/articles/36486-salah-gunung-lewotobi-laki-laki-erupsi-usai-gempa-ntt

Mafindo. (2026d). *[Salah] Kebakaran Hutan Kalimantan Merambat Ke Permukiman Brunei Darussalam*. TurnBackHoax. https://turnbackhoax.id/articles/36582-salah-kebakaran-hutan-kalimantan-merambat-ke-permukiman-brunei-darussalam

Mafindo. (2026e). *[Salah] Tsunami Setinggi 10 Meter Terjang Gunungkidul*. TurnBackHoax. https://turnbackhoax.id/articles/35682-salah-tsunami-setinggi-10-meter-terjang-gunungkidul

Mafindo. (2026f). *[Salah] Video "Tsunami menerjang Kuta, Bali"*. TurnBackHoax. https://turnbackhoax.id/articles/32285-salah-video-tsunami-menerjang-kuta-bali

Marangoz, M. M., & Ensar, F. (2025). The Impact of 6+1 Traits Of Writing Model on Seventh-Grade Students' Fairy Tale Writing Skills and Writing Attitudes: A Mixed Methods Study. *International Journal of Current Educational Studies*, *4*(1), 1–22. https://doi.org/10.46328/ijces.159

Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). "Fake news" Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. *American Behavioral Scientist*, *65*(2), 180–212. https://doi.org/10.1177/0002764219878224

Nuthihar, R., Anshori, D. S., Cahyani, I., & Sastromiharjo, A. (2026). Making Students "safe" Writers with the SAFE Model in The Age of AI. *Discourses on Culture*, *25*(1), 189–220. https://doi.org/10.2478/doc-2026-0008

Pimentel, D. R., & Campbell, V. (2025). Flipping The Script on Claim, Evidence, and Reasoning: The Credibility CER (C-CER). *The Science Teacher*, *92*(4), 37–41. https://doi.org/10.1080/00368555.2025.2500938

Rubin, V. L. (2019). Disinformation and Misinformation Triangle: A Conceptual Model for "Fake News" Epidemic, Causal Factors and Interventions. *Journal of Documentation*, *75*(5), 1013–1034. https://doi.org/10.1108/JD-12-2018-0209

Savira, A. T. D., Nafarin, S. F. A., & Fa'iza, S. N. (2021). Analisis Kesalahan Berbahasa Bidang Morfologi dalam Laporan Isu Hoaks Corona Virus oleh KOMINFO. *GHANCARAN: Jurnal Pendidikan Bahasa Dan Sastra Indonesia*, *3*(1), 1–15. https://doi.org/10.19105/ghancaran.v3i1.4052

Søe, S. O. (2021). A Unified Account of Information, Misinformation, and Disinformation. *Synthese*, *198*(6), 5929–5949. https://doi.org/10.1007/s11229-019-02444-x

T, S. M., & Mathew, S. K. (2022). The Disaster of Misinformation: A Review of Research in Social Media. *International Journal of Data Science and Analytics*, *13*(4), 271–285. https://doi.org/10.1007/s41060-022-00311-6

Talitha, S., Mulyati, Y., Damaianti, V. S., & Kurniawan, K. (2026). Evidence-Based Scaffolding for Digital Multimodal Writing Performance in Blended Learning. *Open Praxis*, *18*(4), 777–794. https://doi.org/10.55982/openpraxis.18.4.1259

Valle, N., Zhao, P., Freed, D., Gorton, K., Chapman, A. B., Shea, A. L., & Bazarova, N. N. (2025). Towards a Critical Framework of Social Media Literacy: A Systematic Literature Review. *Review of Educational Research*, *95*(4), 701–746. https://doi.org/10.3102/00346543241247224

Volkova, S., & Jang, J. Y. (2018). Misleading or Falsification: Inferring Deceptive Strategies and Types in Online News and Social Media. *The Web Conference 2018 - Companion of the World Wide Web Conference, WWW 2018*, 575–583. https://doi.org/10.1145/3184558.3188728

Zubiaga, A., Liakata, M., Procter, R., Wong Sak Hoi, G., & Tolmie, P. (2016). Analysing How People Orient to and Spread Rumours in Social Media By Looking at Conversational Threads. *PLoS ONE*, *11*(3). https://doi.org/10.1371/journal.pone.0150989