

794 **A Qualitative Results**

795 We include qualitative examples from both MONITRS and MONITRS-QA (along with results) in
796 Figure 5.

797 **B Prompts to LLM**

798 We use prompts to LLMs to act as language tools for two types of tasks in our work. The first being to
799 read through and retrieve the relevant information from news articles to caption our image sequences,
800 figures 6 and 7 The second being utilizing our captions to generate event specific question-answer
801 pairs, figures 8 and 9.

802 **C Human Validation of Caption Quality**

803 We conducted human validation on 144 events sampled across 15 disaster types to assess caption
804 quality. Human evaluators were asked to classify each event as: (1) clear alignment between images,
805 captions, and sources, (2) mismatch, or (3) inconclusive where imagery was insufficient to verify
806 caption details. Overall results showed 65.3% clear alignment between images, captions, and sources,
807 18.8% had mismatches, and 16.0% were inconclusive where imagery was insufficient to verify
808 caption details. Excluding inconclusive cases, 77.7% of determinable events showed alignment,
809 demonstrating reasonable caption quality for LLM-generated annotations.

810 Performance varied by disaster type, with strongest results for events with distinct visual signatures.
811 Typhoons, tornadoes, winter storms, and dam-related events achieved 100% accuracy on clear images.
812 Fire events showed 92.3% accuracy (12/13 clear events), coastal storms 90.0% (18/20), and floods
813 85.7% (6/7).

814 Error analysis on mismatched events revealed that snowstorms showed the highest error rates. These
815 errors primarily stem from difficulty distinguishing white snow and ice from clouds or existing snow
816 cover in the imagery. Hurricane events had a 35.7% mismatch rate, largely because captions describe
817 ground-level wind damage that is not visible from satellite perspective.


818 The 16.0% inconclusive rate reflects a persistent challenge in validating satellite based disaster
819 event captions. That is, captions may accurately describe events as reported in news articles, but
820 10m resolution imagery does not provide sufficient detail to verify specific claims. For example,
821 descriptions of "dozens of homes destroyed" cannot be confirmed at this resolution, though large-scale
822 burn scars or flooding extent remain visible. This does not indicate caption errors but rather highlights
823 the resolution gap between textual descriptions from the ground level and satellite imagery. As
824 we discuss in our limitations section 7, we provide complete location and time metadata to enable
825 extensions with higher-resolution data sources.

826 **D International Transfer Evaluation**


827 To assess generalization beyond the United States, we curated a test set of 18 international disaster
828 events from 8 countries across 5 continents: Greece, Chile, Spain, Ecuador, Morocco, Colombia,
829 Libya, Japan, Canada, and Kenya. The set included 5 fires, 3 floods, and 2 earthquakes, with temporal
830 coverage from 2023-2024.

831 We processed these events using our MONITRS pipeline: news article retrieval, location extraction,
832 Sentinel-2 imagery acquisition, and caption generation. For each event, we generated templated
833 multiple-choice questions for event classification, temporal grounding, and location grounding.


834 Our fine-tuned TEOChat achieved 45.65% accuracy averaged across all question types, compared to
835 21.74% for the baseline, TEOchat [14]. On U.S. test data, the fine-tuned model achieved 66.35%
836 versus 26.39% by the baseline. The performance gap suggests that incorporating geographically
837 diverse training data would improve cross-region generalization, though the current results validate
838 that models trained on MONITRS can reasonably generalize to international disasters.




2017-06-15: The Cajete Fire started approximately one mile northeast of Vallecitos de los Indios, burning mostly ponderosa pine.
 2017-06-25: The Cajete Fire, at 1315 acres and 0% contained, continued to spread eastward and southeastward, prompting evacuations of several communities including Ruby Holt Plat, Los Griegos and Sierra de Los Pinos. A community meeting was held the previous evening to inform residents. The Cajete Fire continued to burn, with firefighters working to strengthen containment lines and begin mop-up operations on the north side. Highway 4 remained closed and evacuations were still in effect.
 2017-06-28: The Cajete Fire's progression continued to be monitored, with efforts focused on securing the east and southeast flanks where growth potential remained high. Smoke impacted air quality in the Rio Grande Valley.
 2017-06-30: Firefighters continued to battle the Cajete Fire, focusing on containment and mop-up operations. Hot, dry conditions persisted.
 2017-07-03: The Cajete Fire continued to burn with an impact of smoke on air quality.




2020-08-09: The Grizzly Creek fire grew to 6,251 acres, causing the closure of Interstate 70 between Glenwood Springs and Gypsum, as well as Independence and Cottonwood Passes. Evacuations were ordered for areas east of Glenwood Springs including Lookout Mountain and Coulter Creek.
 2020-08-12: The fire reached the bottom of the drainage. Evacuation orders were lifted for Eagle County residents along Buck Point Drive, though a pre-evacuation order remained in place.
 2020-08-14: Evacuations were ordered for Bair Ranch, Sweetwater, and Coffee Pot Springs; Dolsero was put on pre-evacuation notice. Active fire behavior and Red Flag conditions continued due to gusty winds and low humidity.
 2020-08-17: Firefighters focused on prevention work around the Shoshone Power Plant, Lookout Mountain, and subdivisions. Residents in north Glenwood Springs were warned to prepare for rapidly changing conditions and possible pre-evacuation notices.
 2020-08-19: Hanging Lake was closed due to the fire's proximity. An evacuation center was set up at the Gypsum Recreation Center.



2021-08-14: The Caldor Fire started just east of Omo Ranch and south of Grizzly Flats.
 2021-08-17: The Caldor Fire had burned 6,500 acres by morning and 22,919 acres by 11 p.m.
 2021-08-22: Damage assessment crews reported 104 structures destroyed; an emergency forest closure was issued for the Eldorado National Forest. Damage assessment showed approximately 345 homes destroyed, along with commercial properties and minor structures; firefighters contained about 5% of the fire's perimeter.
 2021-08-24: The Caldor Fire was less than 20 miles from Lake Tahoe; Emerald Bay was shrouded in smoke.
 2021-09-06: Smoke from the Caldor Fire blanketed Lake Tahoe; thousands evacuated South Lake Tahoe due to the fire's proximity.

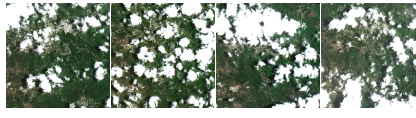


April 12th 2022: The Big Hole Fire began on April 11th, 2022; by April 12th, the fire was actively burning, and one home and 18 outbuildings had already been destroyed.
 May 12th 2022: The Big Hole Fire continued to be actively managed, with crews working on containment lines and rehabilitation efforts. No new significant events are reported between April 14th and this date.




This is a sequence of sentinel-2 satellite images, centered at (29.9096211231, -85.2610062); Comparing satellite images 2 and 5 (assume these show Hurricane Ian's intensification and subsequent impact), which infrastructural damage type shows the most significant change?
 A: Widespread building collapse
 B: Extensive road damage
 C: Major bridge failure
 D: Minimal observable damage

Answers:
 Ours: B
 Gemini: N/A
 TeoChat: A
 Videollava: D




This is a sequence of sentinel-2 satellite images, centered at (18.1127526, -66.2663961); Examining satellite image 3, showing damage from an earthquake?
 A: Residential buildings, showing widespread roof collapses.
 B: Major highways and bridges, showing significant structural damage to multiple crossings.
 C: Agricultural irrigation systems, showing numerous breaks and disruptions.
 D: Power transmission lines, showing widespread outages across the region.

Answers:
 Ours: C
 Gemini: D
 TeoChat: D
 Videollava: D




Q. This is a sequence of sentinel-2 satellite images, centered at (46.7729322, -92.1251218). What natural disaster is occurring in this location?
 a. Volcano
 b. Ice Storm
 c. Fire
 d. Hurricane

Answers:
 Ours: b
 Gemini: d
 TeoChat: d
 Videollava: d



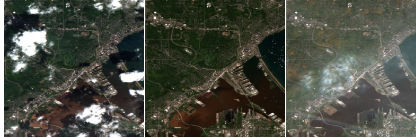
Q. This is a sequence of sentinel-2 satellite images, centered at (35.13458045, -90.05746900). What natural disaster is occurring in this location?
 a. Volcano
 b. Earthquake
 c. Fire
 d. Hurricane

Answers:
 Ours: d
 Gemini: d
 TeoChat: c
 Videollava: c



Q. This is a sequence of sentinel-2 satellite images, centered at (41.9216734, -93.3122705). What natural disaster is occurring in this location?
 a. Severe Storm
 b. Earthquake
 c. Fire
 d. Volcano

Answers:
 Ours: a
 Gemini: b
 TeoChat: b
 Videollava: b



Q. This is a sequence of sentinel-2 satellite images, centered at (46.7729322, -92.1251218). What natural disaster is occurring in this location?
 a. Severe Storm
 b. Earthquake
 c. Fire
 d. Volcano

Answers:
 Ours: a
 Gemini: b
 TeoChat: b
 Videollava: b

Figure 5: Qualitative examples from both MONTRS and MONTRS-QA along with their respective results.

Task: Extract only the event-specific geographical locations mentioned in the provided articles about natural disasters.

Instructions:

1. Carefully review the attached articles about natural disasters and identify **ONLY** proper noun locations that are directly related to where the disaster occurred or had direct impact.
2. Focus on extracting:
 - Specific sites where the event took place (cities, towns, neighborhoods)
 - Precise natural features affected (specific rivers, mountains, forests, beaches)
 - Particular infrastructure impacted (named dams, bridges, parks)
 - Exact regions directly experiencing the disaster effects
3. Present your response in a simple string list format, with each location separated by a comma.
4. If a location appears multiple times, include it only **ONCE** in your list.
5. If the articles contain **NO** specific event locations, return only the word “no” (lower-case).
6. **DO NOT** include:
 - Broad geographical entities not directly affected (countries, states, unless the entire entity was impacted)
 - Locations only mentioned incidentally (headquarters of responding agencies, etc.)
 - Places mentioned for context but not directly experiencing the disaster
 - General areas not specified with proper nouns

Examples:

For a wildfire article: Paradise, Camp Creek Road, Butte County, Sierra Nevada foothills, Eastland County

NOT: California, United States, Western US

For a hurricane article: New Orleans, French Quarter, Lake Pontchartrain, Superdome

NOT: Louisiana, Gulf Coast, United States (unless the entire state/region was directly impacted)

Format for response when locations are found: Paradise, Camp Creek Road, Butte County, Sierra Nevada foothills

Format for response when no locations are found: no

Article Content: {text}

Figure 6: Prompt given to LLM to extract proper nouns locations.

Task: Create a chronological timeline of observable natural disaster events from the provided news articles.

Instructions:

1. Review the attached news articles for information about natural disasters (earthquakes, floods, hurricanes, wildfires, volcanic eruptions, etc.).
2. For each date in the provided list, identify natural disaster events that occurred on or by that date that would be seen remotely.
3. Write a 1-2 sentence description for each date focusing specifically on the visible physical manifestations, such as:
 - Extent of flooding or inundation
 - Wildfire burn scars or active fire fronts
 - Hurricane cloud formations or aftermath flooding
 - Visible structural damage to landscapes or urban areas
 - Changes to coastlines, river courses, or terrain
 - Ash clouds, lava flows, or other volcanic features
4. If a specific date isn't explicitly mentioned in the articles, use context clues to reasonably infer when these visible changes occurred.
5. Present your response as a simple chronological list with dates followed by descriptions.
6. Emphasize the VISUAL aspects that would be detectable from above.

Format example:

June 15, 2023: Extensive flooding covered approximately 60 square miles of the Mississippi Delta region, with standing water clearly visible across previously inhabited areas and farmland.

July 3, 2023: The Caldor wildfire in California created a distinct burn scar spanning 25 miles along the Sierra Nevada mountain range, with active fire fronts visible on the northeastern perimeter.

Article Content: {text}

Dates for analysis: {dates}

Figure 7: Prompt for creating chronological timelines of visually observable natural disaster events

Given a set of statements in an order I'd like you to make 3 multiple choice questions about the events described. Make the questions diverse, covering different aspects of the events that could be answerable using satellite imagery of the event. Each question should have 4 options (A, B, C, and D) with only one correct answer.

Statements: \n{events}

Format your response exactly like this:

****Question 1:**** [Your first question here] A) [First option] B) [Second option] C) [Third option] D) [Fourth option] ****Correct Answer 1:**** [Correct option letter]

****Question 2:**** [Your second question here] A) [First option] B) [Second option] C) [Third option] D) [Fourth option] ****Correct Answer 2:**** [Correct option letter]

****Question 3:**** [Your third question here] A) [First option] B) [Second option] C) [Third option] D) [Fourth option] ****Correct Answer 3:**** [Correct option letter]

Here are some examples of statements: 2021-12-11: No events described in the article are visible from this date. 2021-12-15: Very strong winds in Kansas, Texas, and Oklahoma caused numerous wildfires to spread rapidly. Blowing dust severely reduced visibility, causing streetlights to turn on at midday in some areas. 2021-12-16: A large wildfire in Russell and Ellis Counties, Kansas burned approximately 365,850 acres, destroying at least 10 homes. High winds, gusting up to 100 mph, fueled the fire and other blazes across western Kansas, Oklahoma, and Texas. 2021-12-21: No events described in the article are visible from this date.

Here are some examples of questions:

****Question 1:**** What natural disaster is visible in the satellite images from mid-December 2021? A) Hurricane B) Tornado C) Wildfire D) Flooding ****Correct Answer 1:**** C

****Question 2:**** Approximately how many acres were burned in Russell and Ellis Counties, Kansas? A) 36,585 acres B) 365,850 acres C) 3,658 acres D) 3,658,500 acres ****Correct Answer 2:**** B

****Question 3:**** What weather condition contributed significantly to the spread of wildfires in December 2021? A) Heavy rainfall B) Strong winds C) Freezing temperatures D) High humidity ****Correct Answer 3:**** B

Figure 8: Prompt for generating multiple choice questions from natural disaster event statements

Given a set of statements in an order I'd like you to make 3 questions about the events described. Make the questions diverse, covering different aspects of the events that could be aided answerable using satellite imagery of the event.

Statements: \n{events}

Format your response exactly like this:

****Question 1:**** [Your first question here] ****Answer 1:**** [Your first answer as a complete sentence] ****Question 2:**** [Your second question here] ****Answer 2:**** [Your second answer as a complete sentence]
****Question 3:**** [Your third question here] ****Answer 3:**** [Your third answer as a complete sentence]

Here are some examples of statements: 2021-12-11: No events described in the article are visible from this date. 2021-12-15: Very strong winds in Kansas, Texas, and Oklahoma caused numerous wildfires to spread rapidly. Blowing dust severely reduced visibility, causing streetlights to turn on at midday in some areas. 2021-12-16: A large wildfire in Russell and Ellis Counties, Kansas burned approximately 365,850 acres, destroying at least 10 homes. High winds, gusting up to 100 mph, fueled the fire and other blazes across western Kansas, Oklahoma, and Texas. 2021-12-21: No events described in the article are visible from this date. 2021-12-26: No events described in the article are visible from this date. 2021-12-31: No events described in the article are visible from this date. 2022-01-05: No events described in the article are visible from this date. 2022-01-10: No events described in the article are visible from this date. 2022-01-15: No events described in the article are visible from this date.

Here are some examples of questions:

****Question 1:**** What were the conditions that led to the rapid spread of wildfires in Kansas, Texas, and Oklahoma? ****Answer 1:**** The conditions that led to the rapid spread of wildfires in Kansas, Texas, and Oklahoma were very strong winds, low humidity, and high temperatures.
****Question 2:**** What was the impact of the wildfires in Russell and Ellis Counties, Kansas? ****Answer 2:**** The impact of the wildfires in Russell and Ellis Counties, Kansas was the burning of approximately 365,850 acres and the destruction of at least 10 homes.
****Question 3:**** When did the wildfires in Kansas, Texas, and Oklahoma occur? ****Answer 3:**** The wildfires in Kansas, Texas, and Oklahoma occurred on December 15, 2021.

Figure 9: Prompt for generating question-answer pairs from natural disaster event statements