

# ROUGE : A database of disaster impacts in the Global South using Red Cross reports and Large Language Models

---

HASBINI Laura<sup>\*</sup>, SEVERINO Luca<sup>+\*</sup>, DE BRITO Mariana Madruga, GESUALDO Gabriela, ROTARU Ana Maria, BRESCH David N., MUHLHOFER Evelyn, WANG Jingxian, CARVALHO Tais Maria Nunes

\* corresponding authors: [laura.hasbini@lsce.ipsl.fr](mailto:laura.hasbini@lsce.ipsl.fr) , [luca.severino@usys.ethz.ch](mailto:luca.severino@usys.ethz.ch)

+ presenting author

# Summary

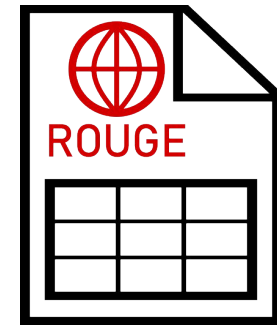
## What?

- ROUGE - a **global database** documenting **socioeconomic impacts of natural hazards**



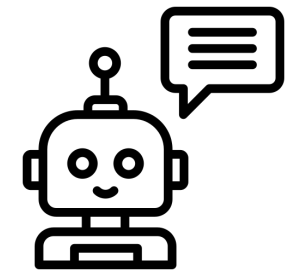
## Why?

- Leverages **unexploited data** and addresses the underrepresentation of the **Global South** in impact databases



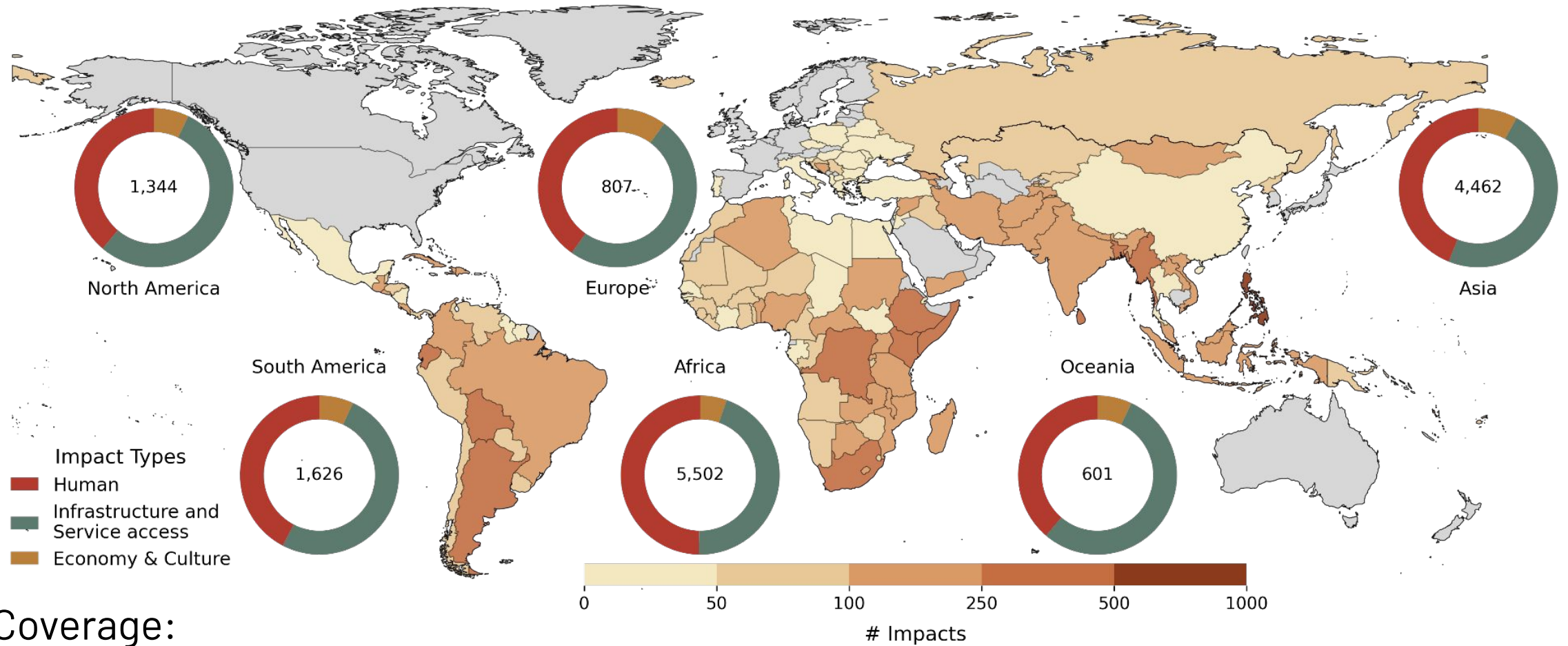
## How?

- Automated data extraction from **Red Cross** reports using **large language models**



LLM

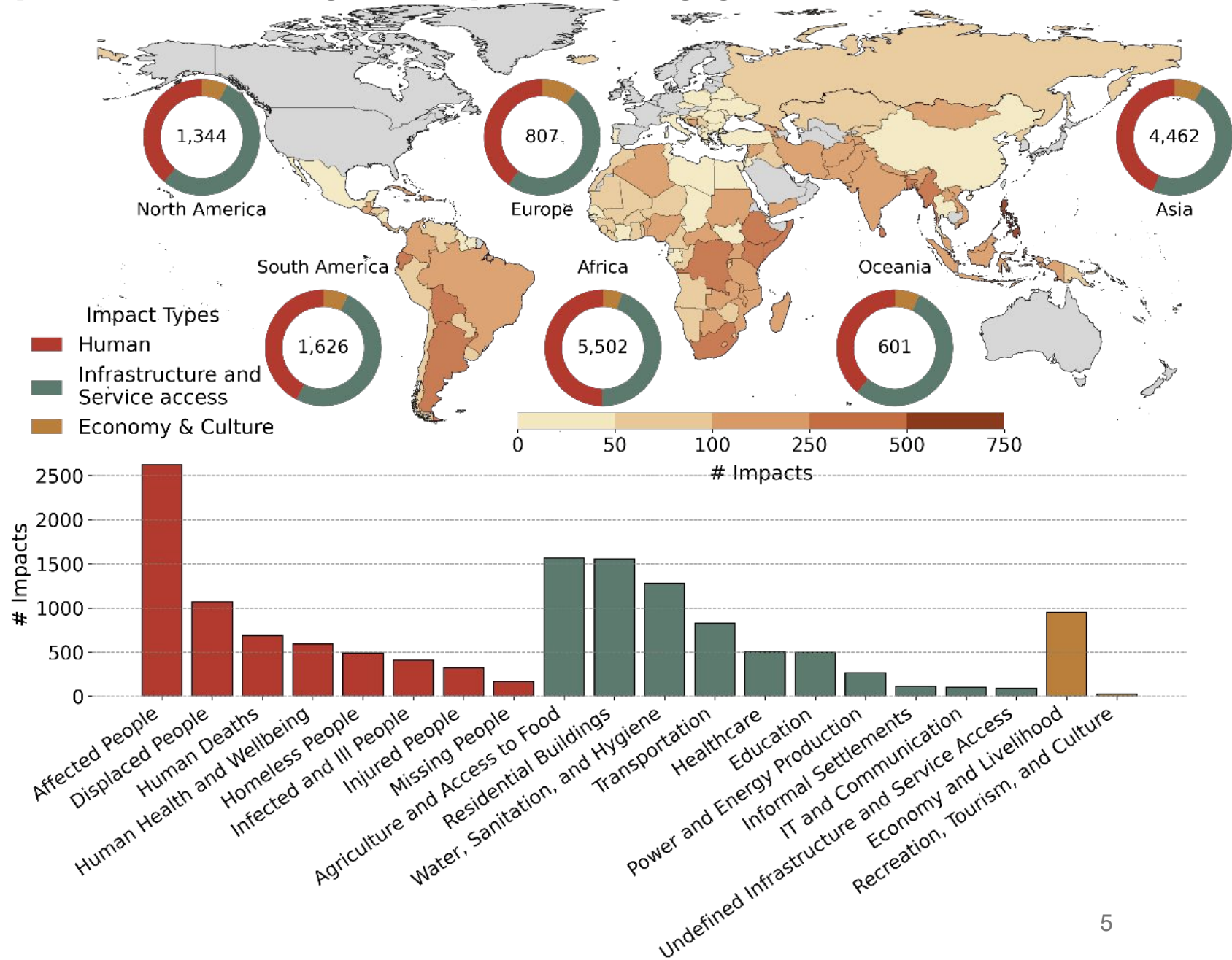
# Overview



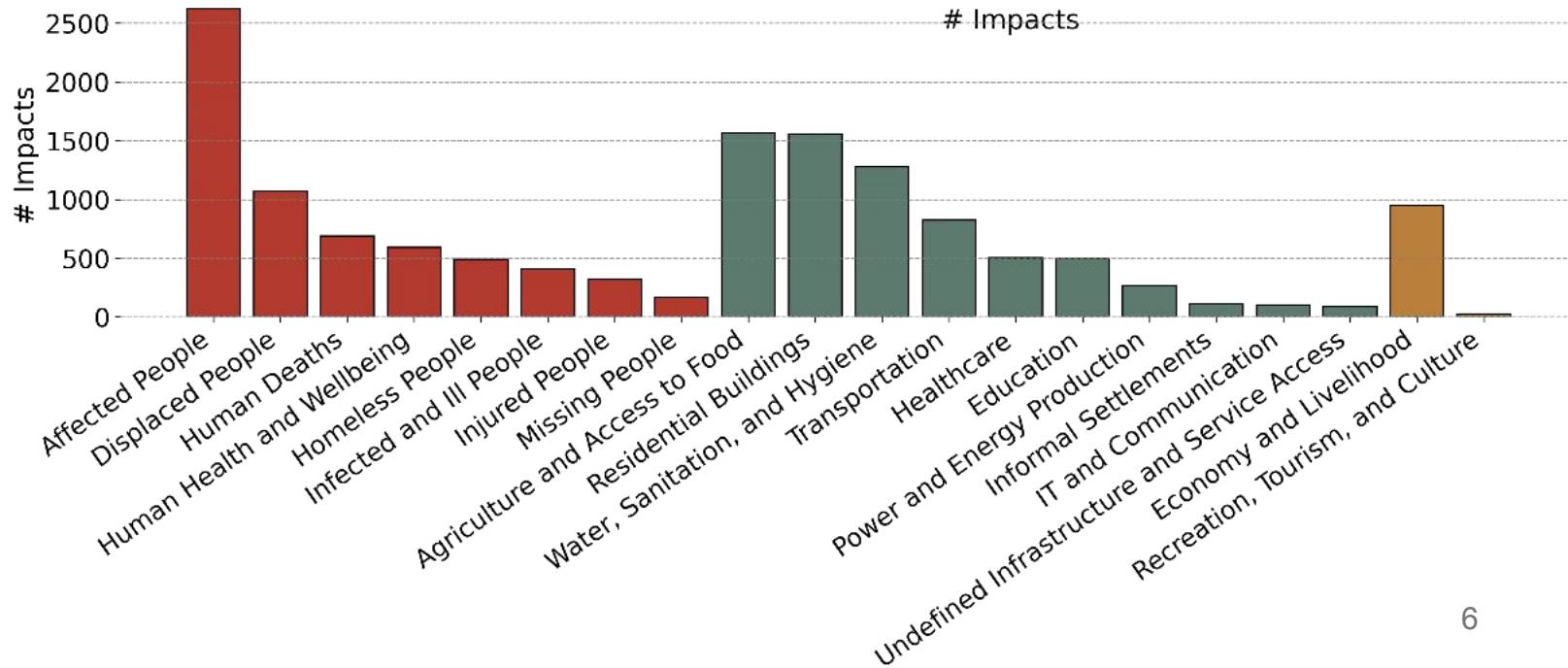
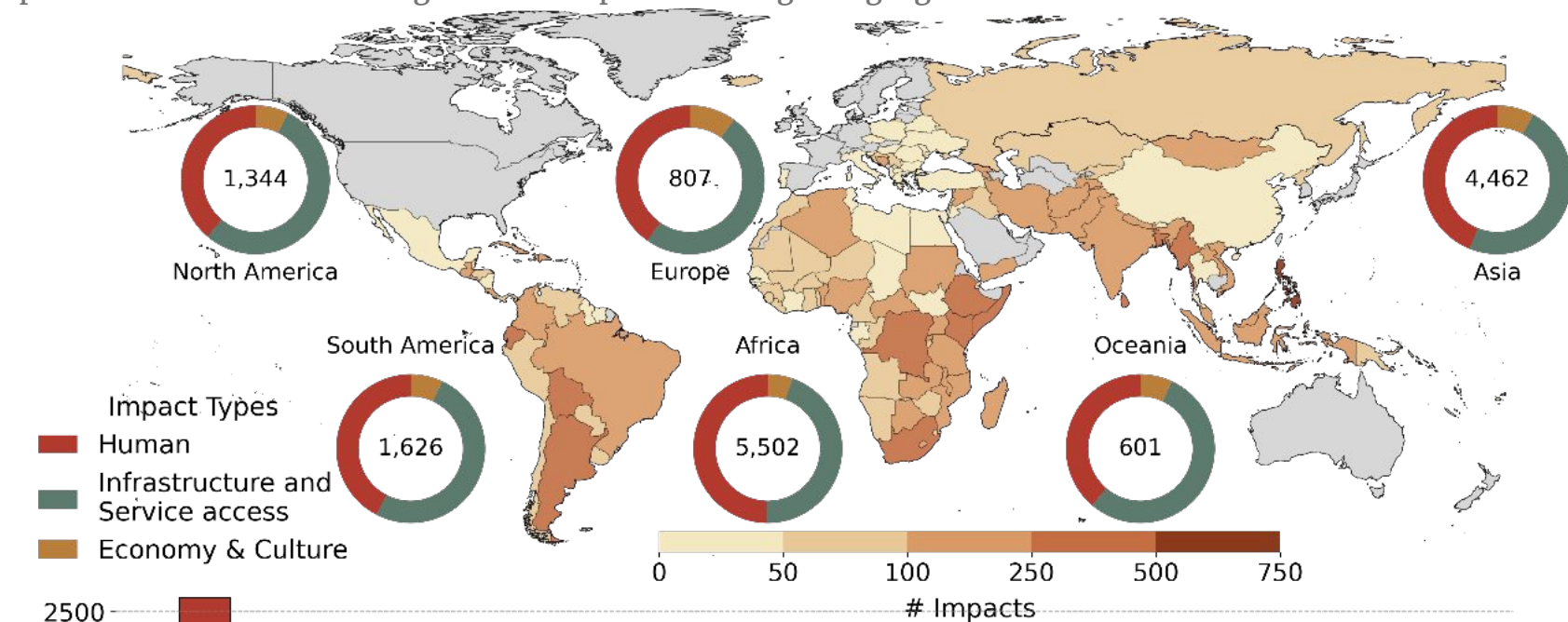
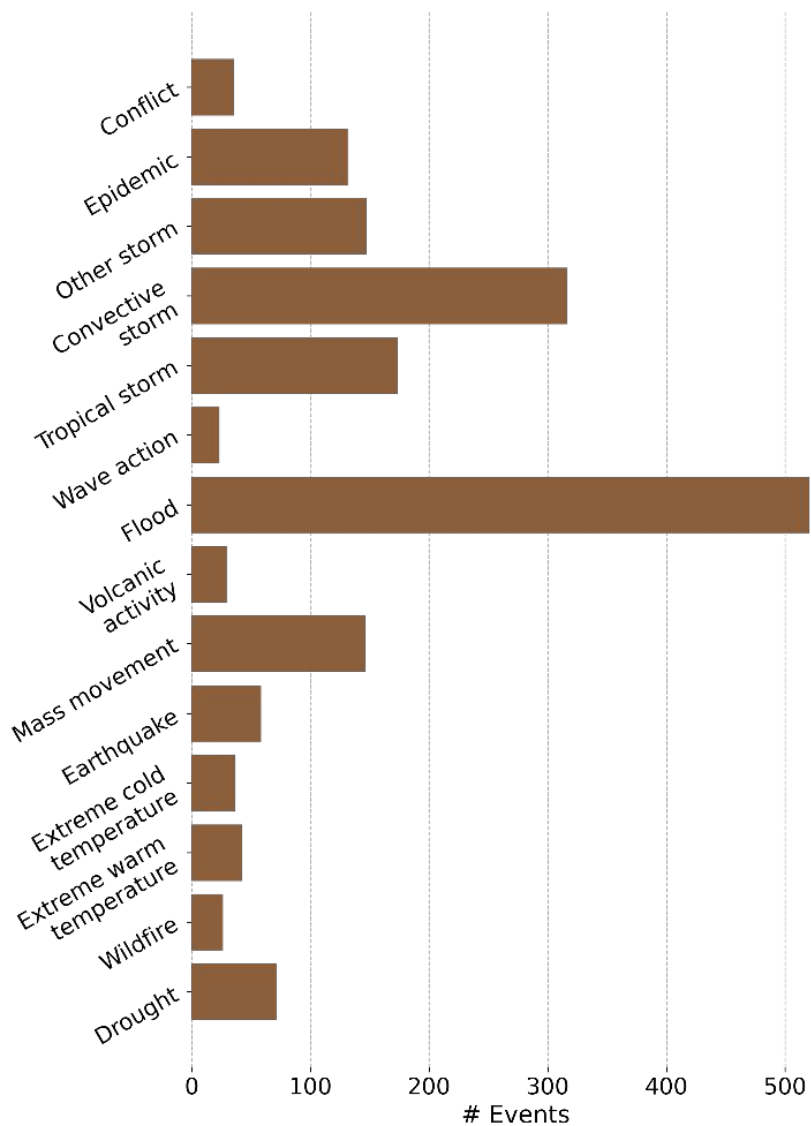
## Coverage:

- 716 disaster reports
- 145 countries
- 2016-2025
- 14138 impacts

# Overview



# Overview



# Input data



## DREF Operation

Sudan Floods 2024

### What happened, where and when?

Sudan has been grappling with heavy rains that have led to widespread flooding across many regions, worsening the already dire situation caused by the conflict that began 16 months ago. Between June 1 and August 12, 2024, DTM Sudan reported 60 incidents of heavy rains and floods, resulting in sudden displacement (IOM). The rainy season is expected to continue until October 2024, with forecasts predicting above-average rainfall. The likelihood of additional flash floods and river flooding remains high. So far, Red Sea, River Nile, and Northern State have been the most severely affected.

Sudan has faced flooding in recent years, but the current humanitarian crisis, combined with heavy rains and deadly floods, has had a devastating impact on communities. Both those displaced by conflict and the host communities supporting them under already challenging conditions are suffering. The floods have rendered roads impassable, further complicating the delivery of aid.

### Scope and Scale

While Sudan is experiencing a conflict which has caused significant damage to its infrastructure, the humanitarian situation in country combined with high level of rains and deadly floods which have led to a devastating impact for both communities, those communities that have been displaced due to the conflict and the host communities which have been supporting those displaced under already dire conditions.

To date, heavy rains, torrents and floods caused 64 loss of life, 1,246 injuries, 39,081 houses are completely destroyed, 11,995 are partially damaged, 270 public facilities and 34,219 latrines are destroyed as well as 10,762 livestock have been perished. The rains destroyed a number of bridges and washed away many main roads, which hindered the movement of goods and passengers in large areas of the affected States. The floods in Northern state and Red Sea are also leading to and increasing displacement.

#### Northern State

In Northern State, the SRCS rapid assessment report indicates that over 50,000 people have been affected, with 10,000 homes completely destroyed and 2,485 partially damaged. Floods and rainwater are besieging large areas of the Northern State, causing the collapse of homes and infrastructure and cutting off the main roads linking the state capital, Dongola, to other cities. Heavy rains that hit the Gold locality in the south of the state led to the complete collapse of 920 houses and the partial collapse of more than 2,000 houses.

## Input

716 operation reports from the International Federation of Red Cross and Red Crescent Societies (IFRC)

# Input data



## DREF Operation

Sudan Floods 2024

### What happened, where and when?

Sudan has been grappling with heavy rains that have led to widespread flooding across many regions, worsening the already dire situation caused by the conflict that began 16 months ago. Between June 1 and August 12, 2024, DTM Sudan reported 60 incidents of heavy rains and floods, resulting in sudden displacement (IOM). The rainy season is expected to continue until October 2024, with forecasts predicting above-average rainfall. The likelihood of additional flash floods and river flooding remains high. So far, Red Sea, River Nile, and Northern State have been the most severely affected.

Sudan has faced flooding in recent years, but the current humanitarian crisis, combined with heavy rains and deadly floods, has had a devastating impact on communities. Both those displaced by conflict and the host communities supporting them under already challenging conditions are suffering. The floods have rendered roads impassable, further complicating the delivery of aid.

### Scope and Scale

While Sudan is experiencing a conflict which has caused significant damage to its infrastructure, the humanitarian situation in country combined with high level of rains and deadly floods which have led to a devastating impact for both communities, those communities that have been displaced due to the conflict and the host communities which have been supporting those displaced under already dire conditions.

To date, heavy rains, torrents and floods caused 64 loss of life, 1,246 injuries, 39,081 houses are completely destroyed, 11,995 are partially damaged, 270 public facilities and 34,219 latrines are destroyed as well as 10,762 livestock have been perished. The rains destroyed a number of bridges and washed away many main roads, which hindered the movement of goods and passengers in large areas of the affected States. The floods in Northern state and Red Sea are also leading to and increasing displacement.

#### Northern State

In Northern State, the SRCS rapid assessment report indicates that over 50,000 people have been affected, with 10,000 homes completely destroyed and 2,485 partially damaged. Floods and rainwater are besieging large areas of the Northern State, causing the collapse of homes and infrastructure and cutting off the main roads linking the state capital, Dongola, to other cities. Heavy rains that hit the Gold locality in the south of the state led to the complete collapse of 920 houses and the partial collapse of more than 2,000 houses.

## Input

716 operation reports from the International Federation of Red Cross and Red Crescent Societies (IFRC)

## Extract

- Impact subtype, magnitude & unit
- Dates
- Locations
- Hazards

# Target database structure



## DREF Operation

Sudan Floods 2024

“ To date, heavy rains, torrents and floods caused 64 loss of life, 1,246 injuries, 39,081 houses are completely destroyed, 11,995 are partially damaged, 270 public facilities and 34,219 latrines are destroyed as well as 10,762 livestock have been perished. The rains destroyed a number of bridges and washed away many main roads, which hindered the movement of goods and passengers in large areas of the affected States. The floods in Northern state and Red Sea are also leading to and increasing displacement ”

## Database structure

<b>appealCode</b>	str	Internal IFRC identifier for each appeal
<b>impactSubtype</b>	str	Category of impact subtype from predefined list
<b>impactValue</b>	float or null	Quantification of impact
<b>impactUnit</b>	str	Unit of impactValue
<b>country</b>	str	String of the countries
<b>location</b>	list	List of locations
<b>geometry</b>	Polygon or MultiPolygon	Polygon of the locations
<b>startDate</b>	date	Start date of impact
<b>endDate</b>	date	End date of impact
<b>hazards</b>	list	List of hazards associated with the described impact

# Target database structure



## DREF Operation

Sudan Floods 2024

“ To date, heavy rains, torrents and floods caused 64 loss of life, 1,246 injuries, 39,081 houses are completely destroyed, 11,995 are partially damaged, 270 public facilities and 34,219 latrines are destroyed as well as 10,762 livestock have been perished. The rains destroyed a number of bridges and washed away many main roads, which hindered the movement of goods and passengers in large areas of the affected States. The floods in Northern state and Red Sea are also leading to and increasing displacement ”

## Quantitative impacts

appealCode	MDRSD034
impactSubtype	Human Deaths
impactValue	64
impactUnit	people
country	Sudan
location	<null>
geometry	<MULTIPOLYGON(...)>
startDate	2024-08-??
endDate	<null>
hazards	[Flood]

# Target database structure



## DREF Operation

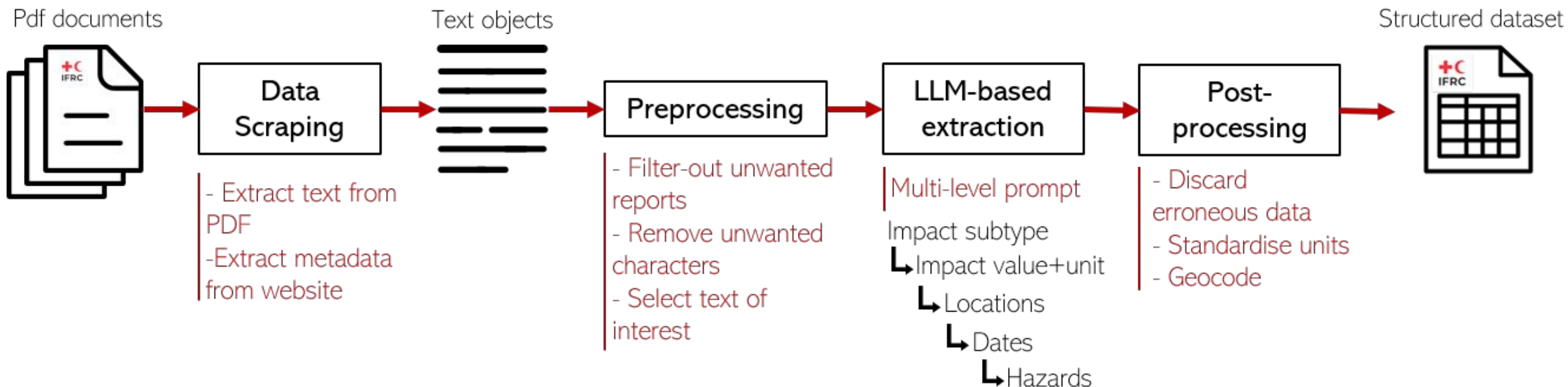
Sudan Floods 2024

“ To date, heavy rains, torrents and floods caused 64 loss of life, 1,246 injuries, 39,081 houses are completely destroyed, 11,995 are partially damaged, 270 public facilities and 34,219 latrines are destroyed as well as 10,762 livestock have been perished. The rains destroyed a number of bridges and washed away many main roads, which hindered the movement of goods and passengers in large areas of the affected States. The floods in Northern state and Red Sea are also leading to and increasing displacement ”

## Qualitative impacts

appealCode	MDRSD034
impactSubtype	Displaced People
impactValue	<null>
impactUnit	people
country	Sudan
location	[Northern State, Red Sea State]
geometry	<MULTIPOLYGON(...)>
startDate	2024-07-??
endDate	<null>
hazards	[Flood]

# Methods

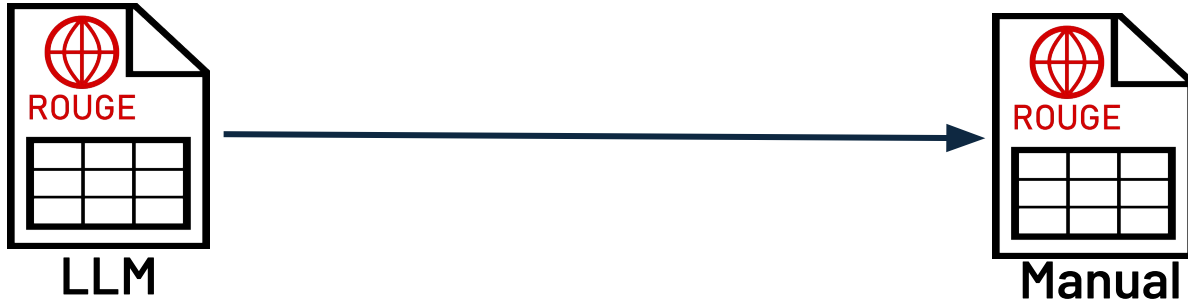


Open source LLM : meta-llama/llama-4-scout-17b-16e-instruct

Open source geocoding : OpenStreetMap Nominatim + Natural Earth + GAUL

# Validation

## Internal Validation



### Event matching

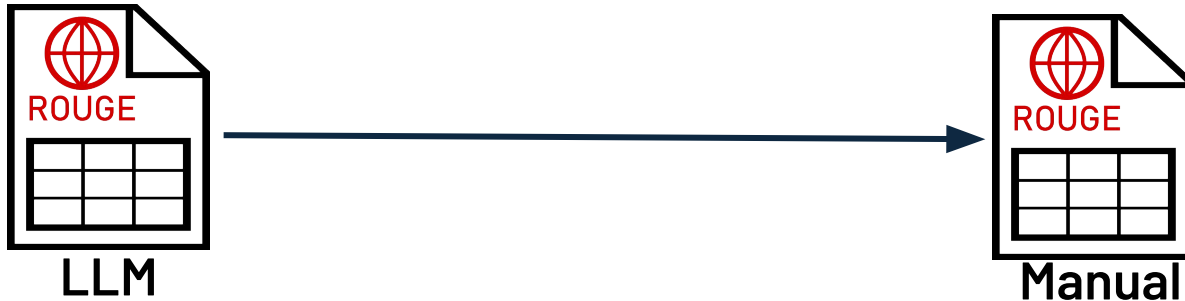
Same appealCode  
By impact line

### Evaluated categories

- All categories

# Validation

## Internal Validation



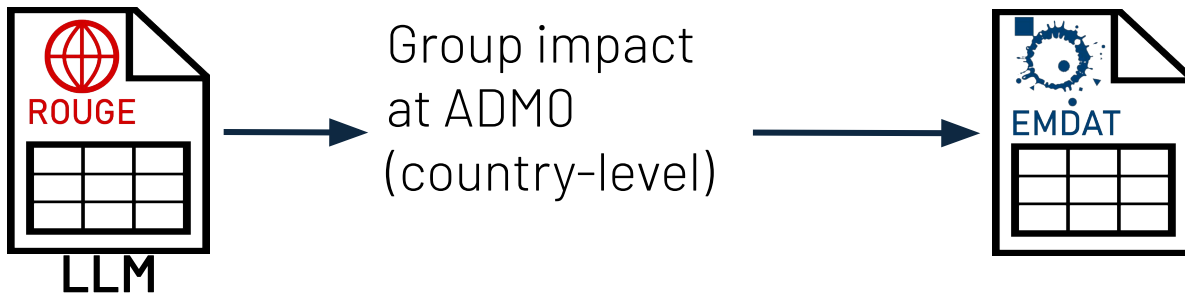
### Event matching

Same appealCode  
By impact line

### Evaluated categories

- All categories

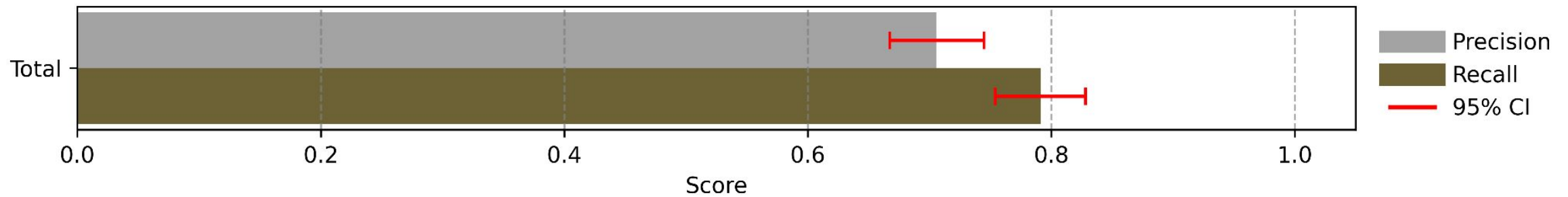
## External Comparison



Same countries  
Same hazards  
Closest event date

- Affected People
- Injured People
- Human Death
- Homeless People

# Internal Validation



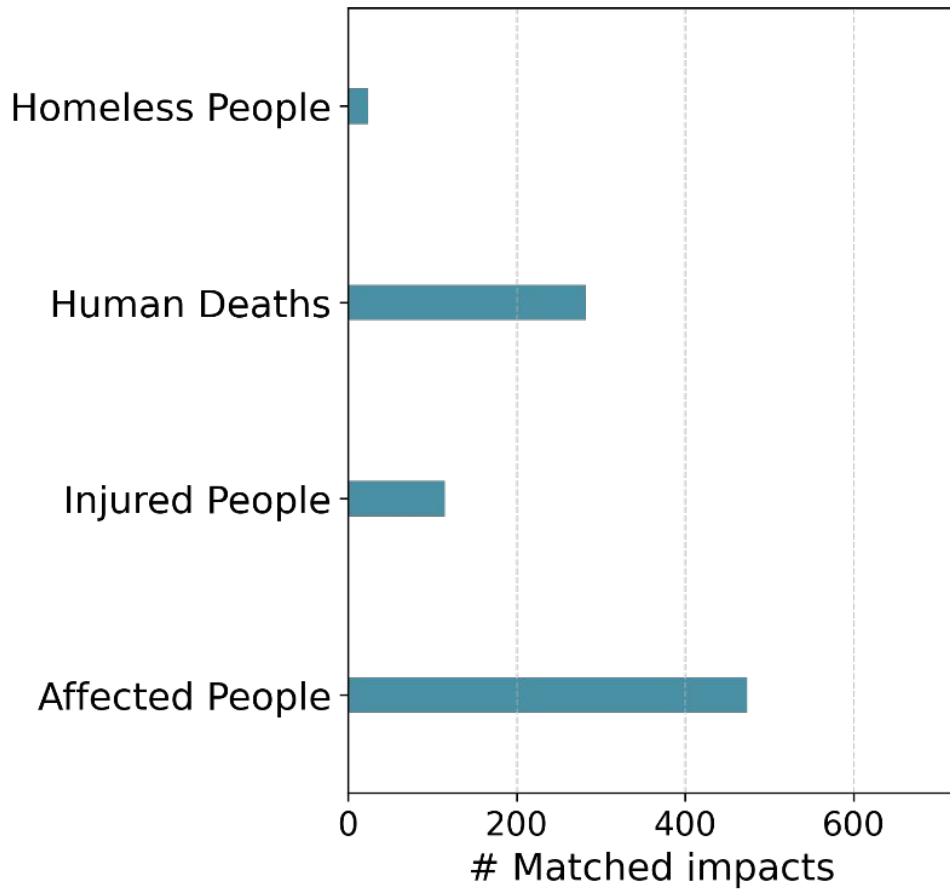
$$\text{precision} = \frac{\text{true positives}}{\text{total extracted}}$$

$$\text{recall} = \frac{\text{true positives}}{\text{total manually labelled}}$$

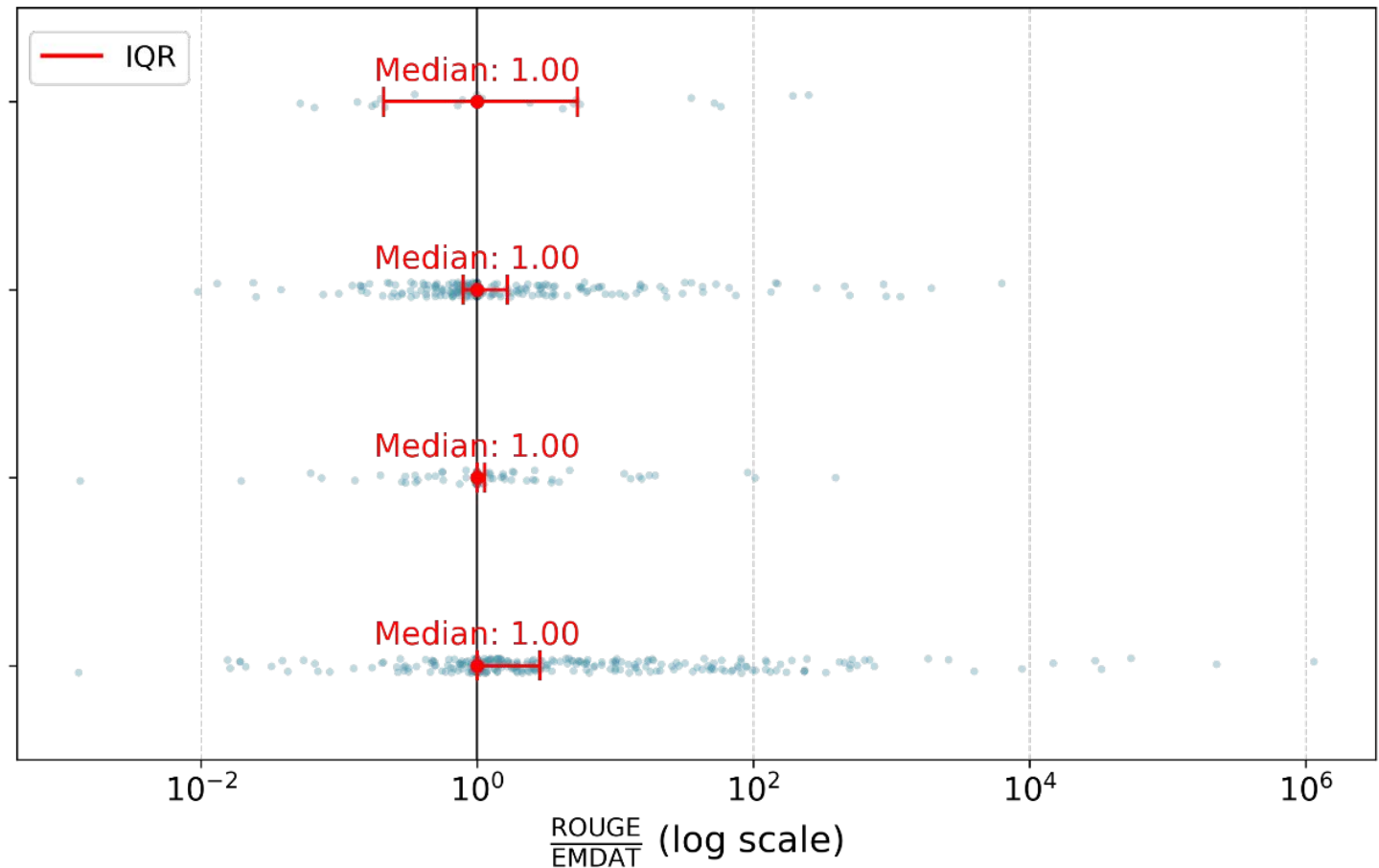
- Overall precision of 0.7, and recall of 0.8

# Comparison with EMDAT

Number of matches



Comparison of impact values (ratio)



# Limitations & Recommendations

## **Current limitations**

- Tables & “dense” text leads to lower recall
- Duplicated mentions of impacts
- General context may be misclassified as impacts

## **Usage recommendations**

- Favor maxing over summing due to duplicates
- Check source excerpts when uncertain
- Review original reports for validation

# Take-aways

## ROUGE

- Database of **human, societal & infrastructural impacts** from **natural hazards**
- **Global** coverage focussing on Global South down to **ADMIN 2** resolution
- Systematic and automatic extraction of **14138 impacts** from 716 disaster reports covering **2016-2025**



Preprint, under revision  
Link to download (preliminary)  
Data in the preprint

## @EGU26

**ITS1.15/NH13.1 - Thursday 16h40 - PICO 4.11**

*Carvalho et al. - Framing impact shaping response : Linking affectedness and action in humanitarian practices*

Corresponding authors:  
[laura.hasbini@lsce.ipsl.fr](mailto:laura.hasbini@lsce.ipsl.fr)  
[luca.severino@usys.ethz.ch](mailto:luca.severino@usys.ethz.ch)

# Climate Risk Workshop 2026

*Benasque Science Center, Spain*

*Sep 20 - Oct 03, 2026*

Interdisciplinary workshop about all things related to climate change, including **advances in climate modelling, impacts on society, ecosystems, and economy, adaptation strategies, and policy development**

## *Confirmed topic highlights*

- climate risk disclosure
- methods in climate justice
- new journal for reproducible climate science
- biodiversity, disturbance, and conservation planning
- inequality as drivers of climate risk
- climate tipping risks
- arts for impact communication

## *More info and registration:*



## **Unconference-style workshop**

—  
flexible, participant-led  
workshop promoting open  
collaboration spaces, diverse  
perspectives, and creative  
dialogues

# Supplementary

# Database structure

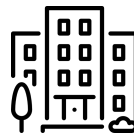
<b>impactSubtype</b>	str	Category of impact subtype from predefined list
<b>impactValue</b>	float	Quantification of impact, unconstrained
<b>impactValueMin</b>	float or None	Lower bound estimate of the impactValue
<b>impactValueMax</b>	float or None	Upper bound estimate of the impactValue
<b>impactUnit</b>	str	Unit of impactValue, constrained
<b>country</b>	str	String of the countries
<b>location</b>	list	List of locations
<b>geometry</b>	Polygon or MultiPolygon	Polygon of the locations
<b>startDate</b>	date	Start date of impact
<b>endDate</b>	date	End date of impact
<b>hazards</b>	list	List of hazards associated with the described impact

## Metadata :

<b>appealCode</b>	str	Appeal code of IFRC report
<b>reportDate</b>	str	Date of the report in YYYY/MM/DD format
<b>reportLink</b>	str	Url link of the pdf report
<b>sourceAnnotation</b>	str	Text excerpts used for the information extraction
<b>nathaz_text</b>	str	Full text used for the information extraction
<b>country_iso3</b>	str	Iso3 of the countries
<b>locationLowestAdmin</b>	str	Lowest administrative level of the row
<b>locationOsm</b>	str	Name of the location found with OpenStreetMap
<b>locationPolygon</b>	str	Name of the location used for the polygon
<b>iso3_code</b>	str	Iso3 of the country found with geocoding

# Impact types

- From topic analysis



Impact Maintype	Impact Subtype
<p style="text-align: center;"><b>Human</b></p>	Affected People
	Injured People
	Displaced People
	Homeless People
	Missing People
	Human Deaths
	Infected and Ill People
	Human Health and Wellbeing
<p style="text-align: center;"><b>Infrastructure and Service access</b></p>	Transportation
	Water, Sanitation and Hygiene
	Healthcare
	IT and Communication
	Residential Buildings
	Informal settlements
	Education
	Power and Energy Production
	Agriculture and Access to Food
Undefined Infrastructure and Service Access	
<p style="text-align: center;"><b>Economy &amp; Culture</b></p>	Recreation, Tourism, and Culture
	Economic and Livelihood

# Hazards types

Inspired by EMDAT types

Hazard	Description
<b>Drought</b>	Prolonged lack of precipitation
<b>Wildfire</b>	Uncontrolled natural fires
<b>Earthquake</b>	Sudden tectonic shifting. Include as well tsunami
<b>Mass movement</b>	Any type of downslope movement of earth materials. Includes Landslides and rockfalls
<b>Volcanic activity</b>	Eruptions and related phenomena
<b>Flood</b>	River, coastal, flash and ice jam flooding
<b>Wave action</b>	Wind-generated surface waves (e.g. Rogue waves, seiche)
<b>Extreme warm temperature</b>	Prolonged, abnormally high heat
<b>Extreme cold temperature</b>	Prolonged, abnormally low cold
<b>Tropical storm</b>	Tropical cyclonic storms (e.g tropical cyclone, tropical storm, hurricanes, typhoons)
<b>Convective storm</b>	Convective-related storms (e.g. severe convective storms, derechos, hail, tornado)
<b>Other storm</b>	Any type of storm which does not correspond to a tropical cyclone or convective storm (e.g. extra-tropical storm, snow storm...)
<b>Epidemic</b>	Sudden outbreak of disease
<b>Conflicts</b>	Disagreements or disputes between different groups, organizations, or states

# Header selection

```
headers_drop = [  
    "coordination and partnerships",  
    "operational strategy",  
    "red cross red crescent action",  
    "operational developments",  
    "summary of response",  
    "summary of the response",  
    "the response so far",  
    "previous operations",  
    "current national society actions",  
    "national society actions",  
    "summary of measures taken by the national society",  
    "detailed operation plan",  
    "summary of the current response",  
    "Overview of Operating National Society Response Action"  
]
```

```
headers_keep = [  
    "operation summary",  
    "situation analysis",  
    "description of the crisis",  
    "what happened, where and when",  
    "description of the disaster",  
    "description of the event",  
    "needs (gaps) identified"  
]
```

# Prompts - Impact Type

**Query:** Identify subtypes of impacts that occurred in the report area described in the text above, only accounting for subtypes described below:

```
{format_desc(impact_desc)}
```

## JSON format:

```
{  
  "impactSubtypes": ["<one or more of {list(impact_desc.keys())}>" or null], # Description: The list of  
  subtypes of impact (e.g., Affected People, Road Infrastructure, Crop Production and Forestry).  
}
```

## Rules

- Only include impact subtypes that are in the following list: {list(impact\_desc.keys())}.
- Follow the JSON format strictly; do not add or remove fields.
- Ensure all field constraints on allowed values and their data types are respected.
- Do not add notes or extra text, only output the JSON

## Example output:

```
{"impactSubtypes" : ["Affected People", "Crop Production and Forestry", "Healthcare Infrastructure"]}
```

# Prompts – Value Unit

**Query:** Based on the list of types of impacts that you previously identified: {answer} identify the values and units of these impacts from the text above.

## **JSON format:**

```
[ { {
```

```
"impactSubtype": "<one of {answer}>", # Description: Subtype of impact (e.g., Affected People, Road Infrastructure, Crop Production and Forestry).
```

```
"impactValue": <float or null> # Description: The quantified value of the impact. Provide the exact number if mentioned. If a range is provided give the upper estimate. Use null if unknown.
```

```
"impactValuePrecision": "one of ["exact", "approx"]", # Description: The flag describing whether the quantified impact value is exact or approximate. Use null if unknown.
```

```
"impactValueMin": <float or null>, # Description: The lower bound estimate of the quantified value of the impact if the impact is approximate or a range. Use null if unknown.
```

```
"impactValueMax": <float or null>, # Description: The upper bound estimate of the quantified value of the impact if the impact is approximate or a range. Use null if unknown.
```

```
"impactUnit": "<string>" or null, # Description: The unit of the impact value (e.g., people, meters, houses). Use null if unknown.
```

```
"valueAnnotation": ["<list of strings>"] # Description: the exact text excerpt from where you extracted the impacts information. Write the text exactly as found in the original text.}}]
```

# Prompts – Value Unit

## Rules

- The `impactSubtype` field must only contain one of the valid values: {answer}.
- Do not reuse a specific impactValue for different entries.
- Provide the exact text excerpt from where you extracted the impacts information in the `valueAnnotation` field. Write the text exactly as found in the original text. Provide the entire sentences.
- Extract each mention of an impact only once. Do not repeat yourself.
- Provide the unit of the impact in the `impactUnit` field exactly as found in the original text, keeping all information on the measured quantity (e.g. write 'km of roads' instead of just 'km')
- When multiple numbers are provided favour the `impactValue` associated with `impactUnit` in terms of "people" over "households", and "CHF" over other currencies
- When the `impactValue` field is an exact number, `impactValuePrecision` must be set to `exact`.
- When the `impactValue` field is an approximation, `impactValuePrecision` must be set to `approx`.
- When the `impactValue` field is a range, `impactValuePrecision` must be set to `approx` and the `impactValueMin` and `impactValueMax` fields must be filled.
- Follow the JSON format strictly; do not add or remove fields.
- Ensure all field constraints on allowed values and their data types are respected.
- Do not add notes or extra text, only output the list of JSONs.

# Prompts – Value Unit

**Example output:**""[

```
{  "impactSubtype" : "Crop Production and Forestry",  
  "impactValue":null,  
  "impactUnit": "kg of crop production",  
  "impactValuePrecision" : "approx",  
  "impactValueMin": 100,  
  "impactValueMax": 200,
```

```
  "valueAnnotation": ["Flash floods and tsunamis impacted 100 to 200 kg of crop production in Red Sea State  
between August to October 2024"]},
```

```
{  "impactSubtype" : "Healthcare Infrastructure",  
  "impactValue": 4,  
  "impactUnit": "healthcare facilities",  
  "impactValuePrecision" : "approx",  
  "impactValueMin": 4,  
  "impactValueMax": null,
```

```
  "valueAnnotation": ["At least 4 hospitals have been impacted by a hailstorm in River Nile State alone." ]}]"" 31
```

# Prompts - Location

**Query:** Identify the locations where {impact\_description} occurred in the report area described in the text above.

**JSON format:**

```
{  
  "country": ["<list of strings>"], # Description: The list of affected countries where the described  
  impact occurred.  
  "location" : ["<list of strings>"] or null, # Description: The list of affected locations at the  
  subnational level (e.g. cities, regions) where the described impact occurred.  
  "locationAnnotation": ["<list of strings>"] # Description: the exact text excerpt from where you  
  extracted the location information. Write the text exactly as found in the original text.  
}
```

# Prompts - Location

## Rules

- Provide the the most precise level of location found for the described impact in the `city field`.
- Give the location names exactly as written in the text. Do not aggregate different locations in words such as "various locations" but provide the list of names.
- Follow the JSON format strictly; do not add or remove fields.
- Provide the exact text excerpt from where you extracted the location information in the `locationAnnotation` field. Write the text exactly as found in the original text. Provide the entire sentences.
- Follow the JSON format strictly; do not add or remove fields.
- Ensure all field constraints on allowed values and their data types are respected.
- Do not add notes or extra text, only output the JSON.

## Example output:

```
{  "country" : ["Sudan"],
  "location" : ["Abu Hamad", "Tokar"]
  "locationAnnotation": ["Landslides impacted 10000 people in the cities of Abu Hamad and Tokar on the 29 August 2024"]}
```

# Prompts - Date

**Query:** Identify the dates when {impact\_description} occurred at {locations} in the report area described in the text above.

**JSON format:**

```
{
  "startYear": <integer or null>, # Description: The year when the impact started. Use null if
unknown.
  "startMonth": <integer or null>, # Description: The month when the impact started. Use null if
unknown.
  "startDay": <integer or null>, # Description: The day when the impact started. Use null if unknown.
  "endYear": <integer or null>, # Description: The year when the impact ended. Use null if ongoing or
unknown.
  "endMonth": <integer or null>, # Description: The month when the impact ended. Use null if
ongoing or unknown.
  "endDay": <integer or null>, # Description: The day when the impact ended. Use null if ongoing or
unknown.
  "dateAnnotation": ["<list of strings>"] # Description: the exact text excerpt from where you
extracted the date information. Write the text exactly as found in the original text.
}
```

# Prompts - Date

## Rules

- Follow the JSON format strictly; do not add or remove fields.
- If you cannot disentangle between different dates, give the earliest date for the starting date and the latest date for the ending date.
- Provide the exact text excerpt from where you extracted the date information in the `dateAnnotation` field. Write the text exactly as found in the original text. Provide the entire sentences.
- Ensure all field constraints on allowed values and their data types are respected.
- Do not add notes or extra text, only output the JSON.

## Example output:

```
{  "startYear" : 2024,  
  "startMonth" : 08,  
  "startDay" : 29,  
  "endYear" : null,  
  "endMonth" : null,  
  "endDay" : null,  
  "dateAnnotation": ["Landslides impacted 10000 people in the cities of Abu Hamad and Tokar on the 2935  
August 2024"]}]
```

# Prompts - Hazard

**Query:** Identify the hazards that caused {impact\_description} at {locations} at {dates} in the report area described in the text above.

## **JSON format:**

{ "hazards": ["<one or more of {hazards\_list}>" or null], # Description: List of hazards causing the impact (e.g., Tropical storm, Drought, Flood).

"hazardsAnnotation": ["<list of strings>"] # Description: the exact text excerpt from where you extracted the hazard information. Write the text exactly as found in the original text. Provide the entire sentences. }

# Prompts - Hazard

## Rules

- The `hazards` field must only contain values from {hazards\_list}.
- Provide the exact text excerpt from where you extracted the hazard information in the `hazardsAnnotation` field. Write the text exactly as found in the original text.
- Follow the JSON format strictly; do not add or remove fields.
- Ensure all field constraints on allowed values and their data types are respected.
- Do not add notes or extra text, only output the JSON.

## Example output:

```
{"hazards" : ["Mass movement"]
```

```
"hazardsAnnotation": ["Landslides impacted 10000 people in the cities of Abu Hamad and Tokar on the 29 August 2024"]}]
```

```
{"hazards" : ["Convective storm"]
```

```
"hazardsAnnotation": ["Hailstorms impacted 100 to 200 kg of crop production in Red Sea State between August to October 2024"]}]
```

# Prompts - Hazard

Using information from the text above and no previous knowledge, please answer the query.

**Query:** Identify subtypes of impacts that occurred in the report area described in the text above, only accounting for subtypes described below:

```
{format_desc(impact_desc)}
```

Answer as a JSON in the following format and respecting the rules described after:

JSON format:

```
{  
  "impactSubtypes": ["<one or more of {list(impact_desc.keys())}>" or null], # Description: The list of  
  subtypes of impact (e.g., Affected People, Road Infrastructure, Crop Production and Forestry).  
}
```

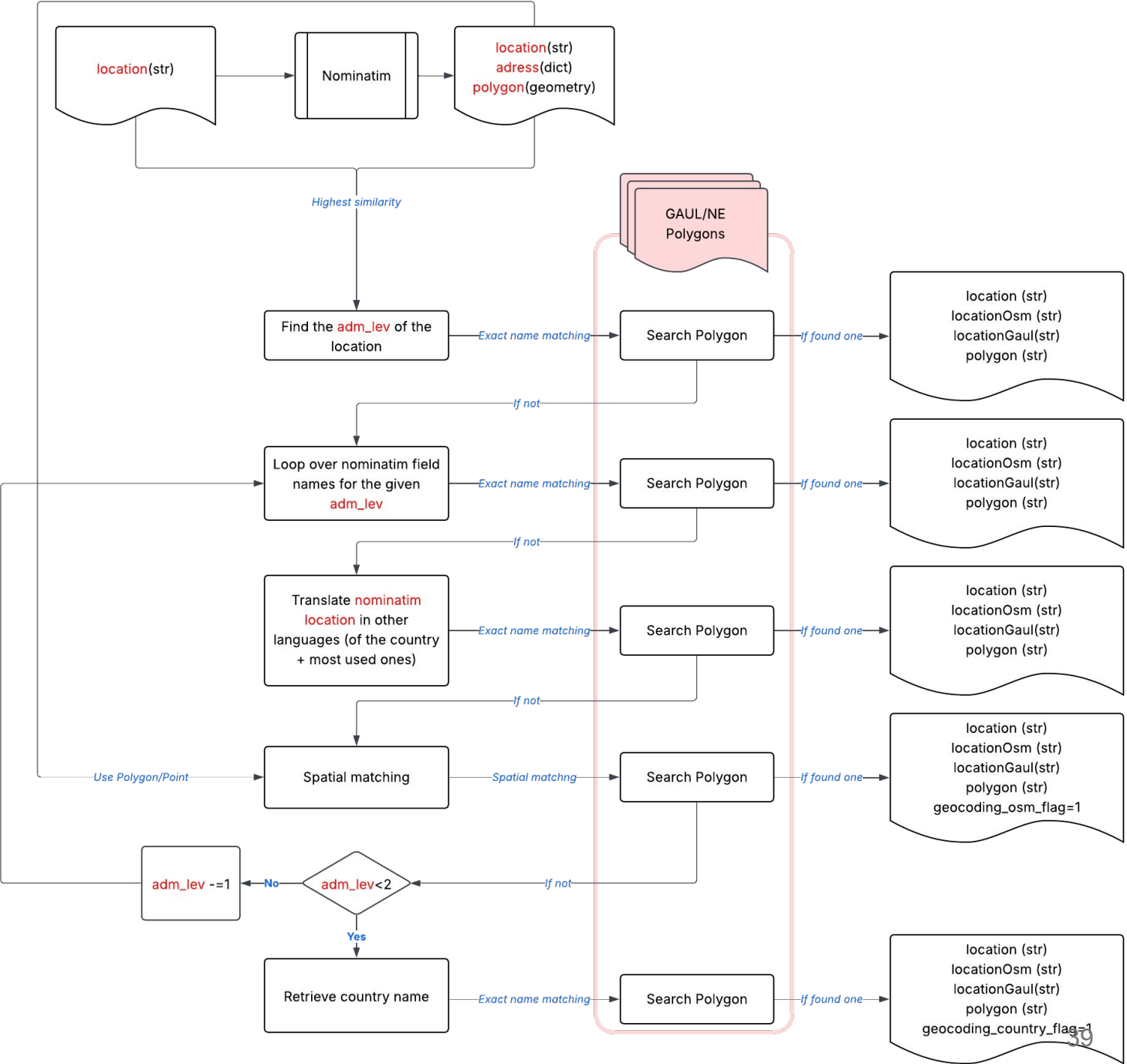
## Rules

- Only include impact subtypes that are in the following list: {list(impact\_desc.keys())}.
- Follow the JSON format strictly; do not add or remove fields.
- Ensure all field constraints on allowed values and their data types are respected.
- Do not add notes or extra text, only output the JSON

## Example output:

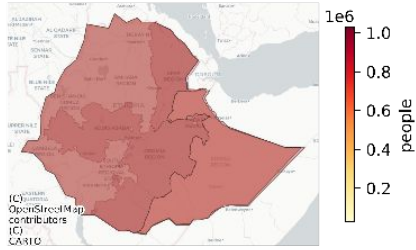
```
{"impactSubtypes" : ["Affected People", "Crop Production and Forestry", "Healthcare Infrastructure"]}
```

# Geocoding

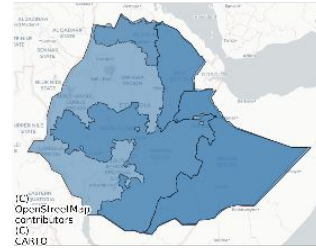


# Impact Distribution for Appeal: MDRET034 ['ETH'], ['2025-02-25']

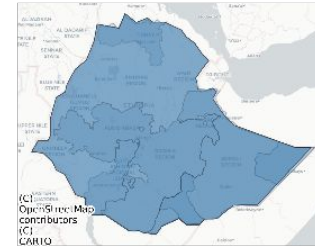
**Affected People**  
(6 records, Max: 1,030,162)



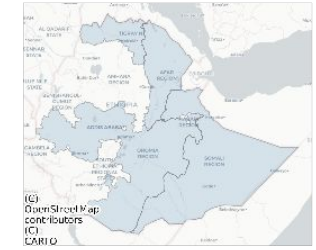
**Agriculture and Access to Food**  
(10 records, Max: 30,000)



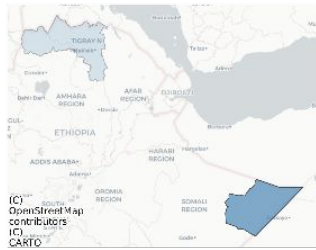
**Displaced People**  
(7 records, Max: 450,000)



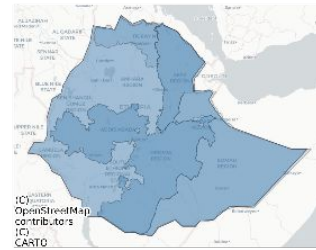
**Economy and Livelihood**  
(1 records)



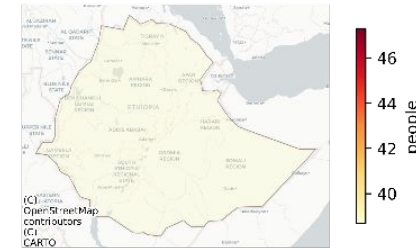
**Education**  
(6 records, Max: 5,332)



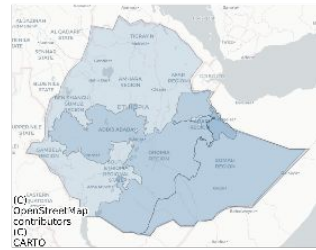
**Healthcare**  
(5 records, Max: 13,000)



**Human Deaths**  
(3 records, Max: 43.0)



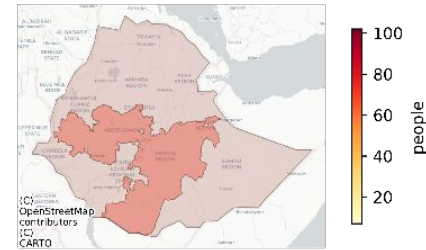
**Human Health and Wellbeing**  
(1 records)



**IT and Communication**  
(1 records)



**Infected and Ill People**  
(5 records, Max: 102.0)



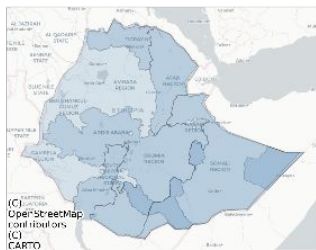
**Power and Energy Production**  
(1 records)



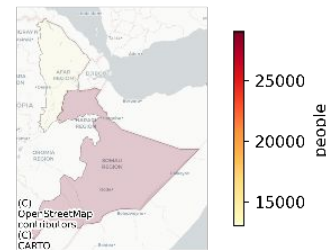
**Transportation**  
(2 records, Max: 1.0)



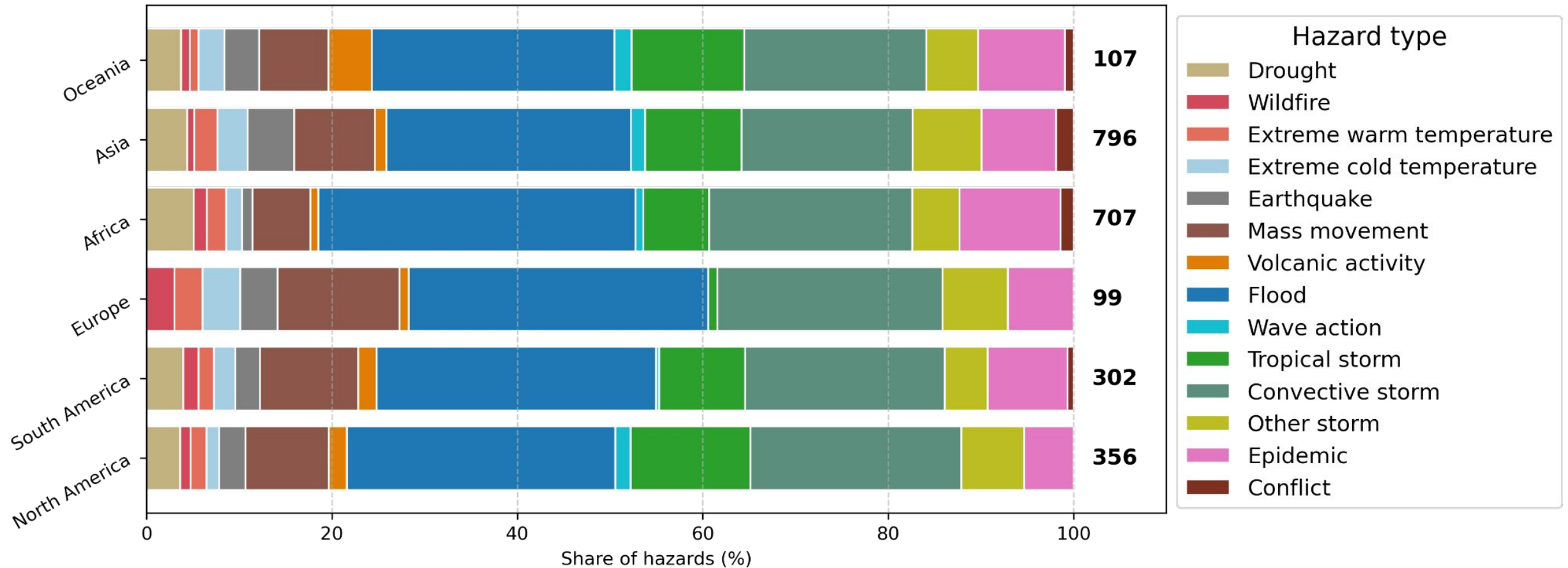
**Undefined Infrastructure and Service Access**  
(1 records)



**Water, Sanitation, and Hygiene**  
(5 records, Max: 29,046)



# Hazard overview



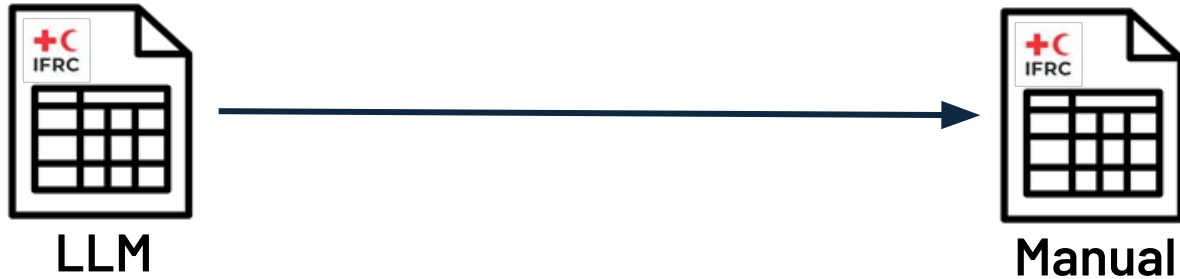
- Most damage associated to floods and convective storm events
- Non-negligible proportion of Epidemics and Conflicts → Compound events

# Output example

appeal...	reportDate	impactSubtype	# impactValue	# impactValueMin	# impactValueMax	impactUnit	locationOsm	co...	hazards
MDRSD034	2025-03-17	Affected People	460743.0	Missing value	Missing value	people	['Kassala', 'Sudan', 'Nile', 'Northern S	['Sudan']	['Flood', 'Epidemics', 'Conflict']
MDRSD034	2025-03-17	Injured People	1246.0	Missing value	Missing value	people	['Sudan']	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Displaced People	178460.0	Missing value	Missing value	people	['Sudan', 'River Nile State']	['Sudan']	['Flood']
MDRSD034	2025-03-17	Homeless People	185805.0	Missing value	Missing value	people	['Kassala', 'Red Sea State', 'West Darf	['Sudan']	['Flood', 'Epidemics', 'Conflict']
MDRSD034	2025-03-17	Human Deaths	64.0	Missing value	Missing value	people	['Sudan']	['Sudan']	['Flood', 'Epidemics', 'Conflict']
MDRSD034	2025-03-17	Infected and Ill People	57000.0	1000.0	1000.0	cases	['Kassala', 'Nile', 'White Nile', 'Al Gazi	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Residential Buildings	39081.0	Missing value	Missing value	homes	['Kassala', 'West Darfur', 'White Nile',	['Sudan']	['Flood', 'Conflict', 'Epidemics']
MDRSD034	2025-03-17	Water, Sanitation, and Hygiene Infra:	1380.0	Missing value	Missing value	water, sanitation and hygiene faciliti	['Port Sudan', 'Tokar', 'Sudan']	['Sudan']	['Flood']
MDRSD034	2025-03-17	Healthcare Infrastructure	80.0	Missing value	Missing value	% of healthcare facilities	['Sudan']	['Sudan']	['Conflict', 'Flood', 'Epidemics']
MDRSD034	2025-03-17	Road Infrastructure	Missing value	Missing value	Missing value	Missing value	['Northern State', 'Red Sea State', 'To	['Sudan']	['Flood']
MDRSD034	2025-03-17	Affected Livestock and Animals	10000.0	Missing value	Missing value	affected animals	['Khawr Arba`at']	['Sudan']	['Flood']
MDRSD034	2025-03-17	Other Infrastructural Impacts	66.359	Missing value	Missing value	homes	['Kassala', 'Red Sea State', 'West Darf	['Sudan']	['Flood', 'Conflict', 'Epidemics']
MDRSD034	2025-03-17	Informal settlements	Missing value	Missing value	Missing value	Missing value	['Sudan']	['Sudan']	['Flood', 'Epidemics', 'Conflict']
MDRSD034	2025-03-17	Education Infrastructure	70.0	Missing value	Missing value	education facilities	['Port Sudan']	['Sudan']	['Flood']

# Validation

## Internal



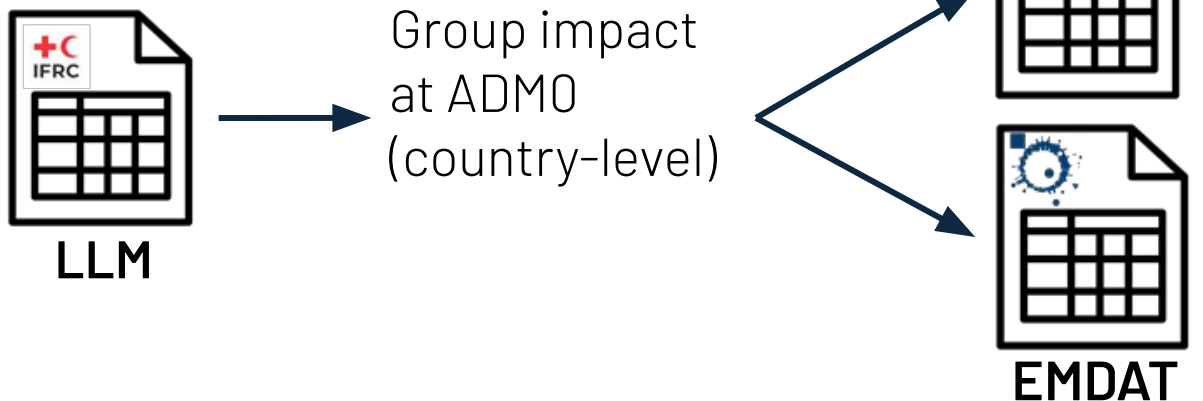
### Event matching

Same appealCode  
By impact line

### Evaluated categories

- All categories

## External



Same appealCode

- Affected people
- Injured people
- Displaced People
- Human Deaths
- Missing People

Same countries  
Same hazards  
Closest event date

- Affected People
- Human Death
- Homeless People

# Internal Validation

Computed in number of rows

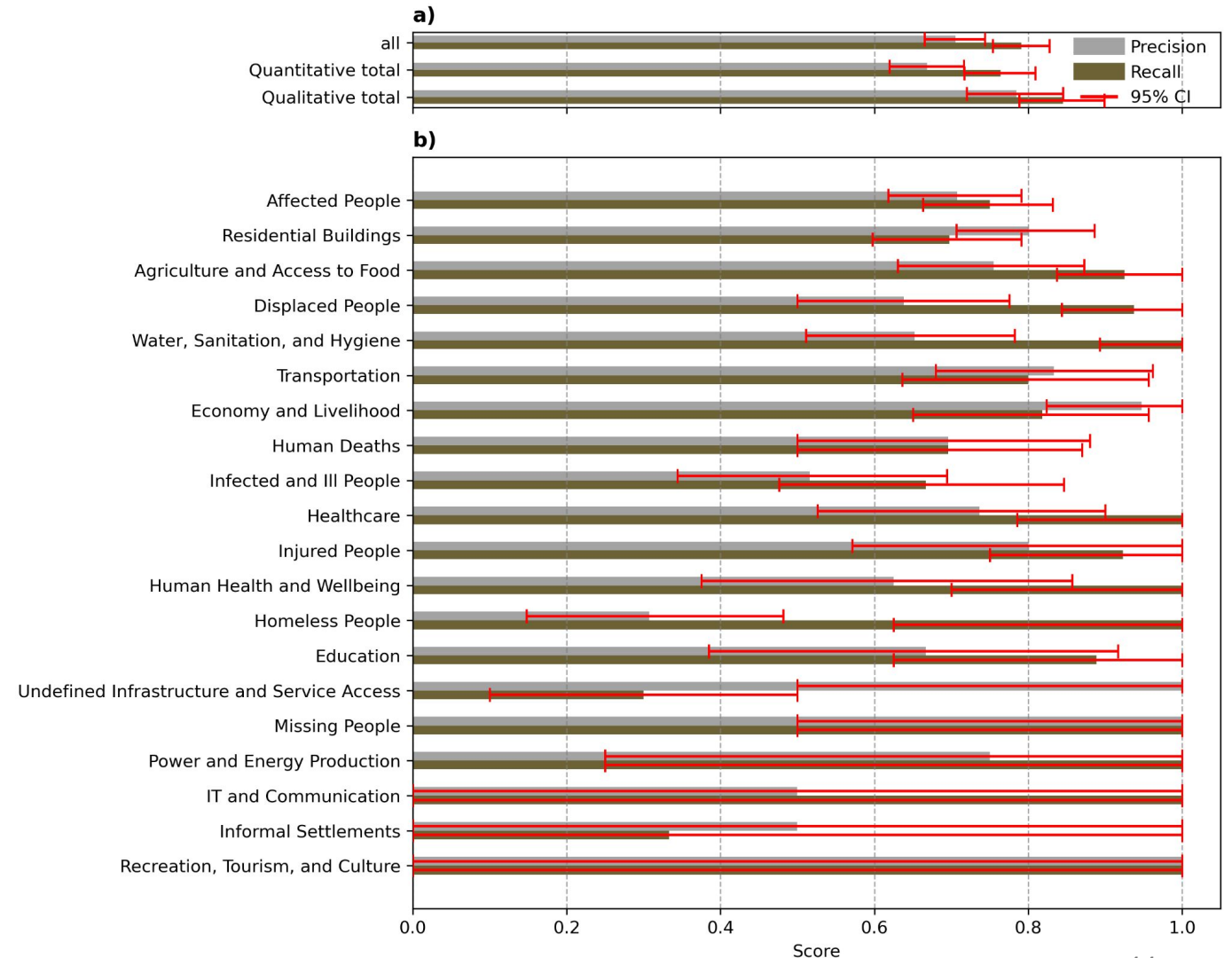
True Positive (**TP**):  
LLM matches with Manually

False Positive (**FP**):  
LLM rows unmatched with Manual

False Negative (**FN**):  
Manual rows unmatched with LLM

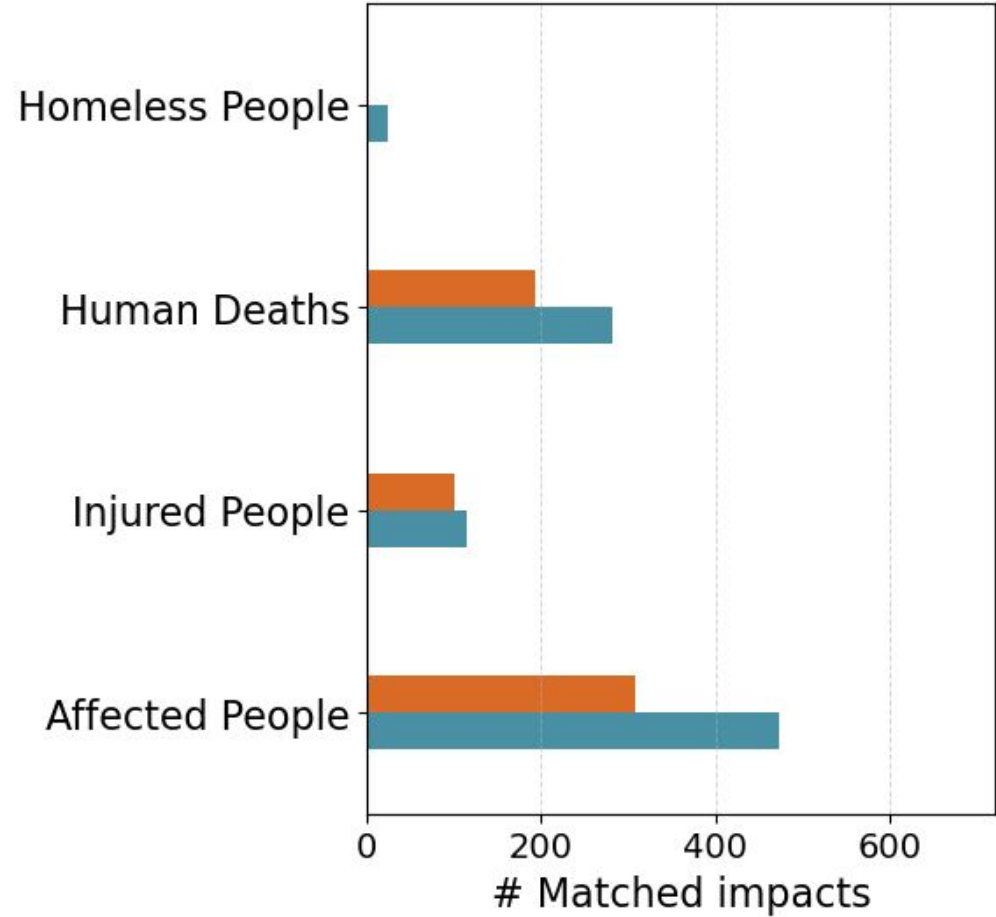
$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

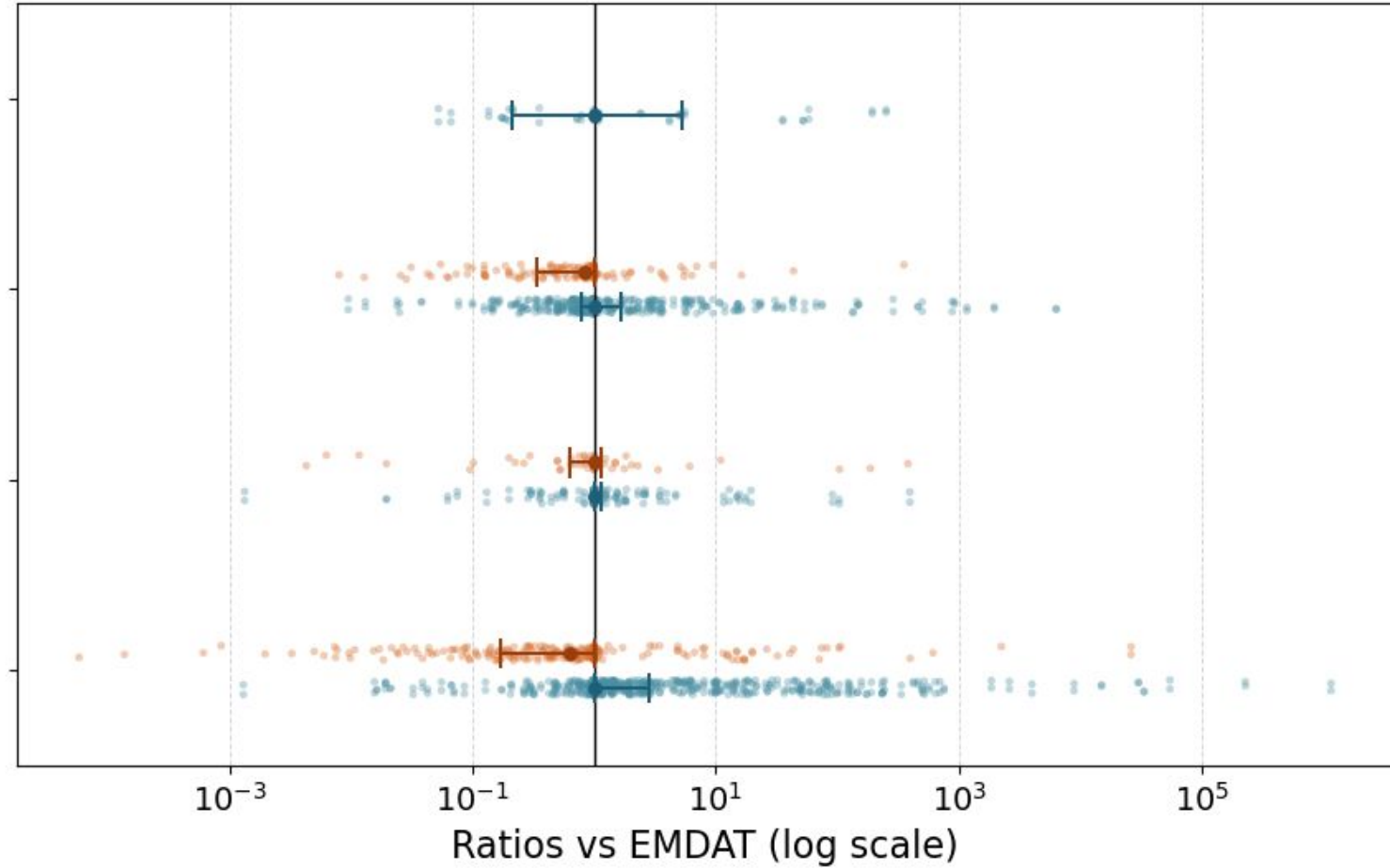


# External Validation

(a)



(b)



● ROUGE/EMDAT ● IFRCGO/EMDAT — IQR

# Similarity per variable

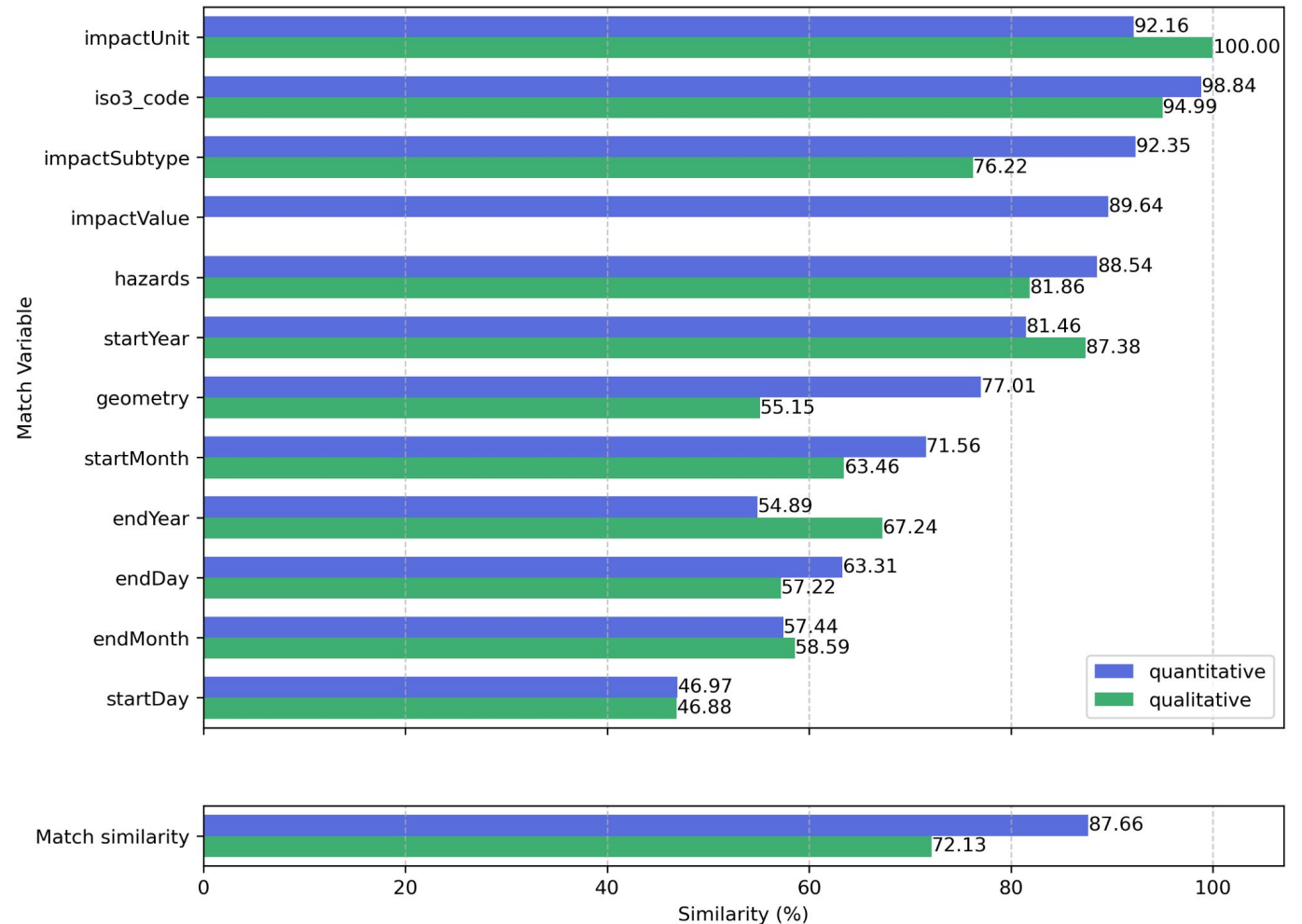
Intersection-Over-Union  
(geometry)

$$IoT(x, y) = \frac{\cap(x, y)}{\cup(x, y)}$$

Value Similarity  
(impactValue)

$$impactValueSim = 1 - \max\left(\left|\frac{v_1 - v_2}{v_1}\right|, \left|\frac{v_2 - v_1}{v_2}\right|\right)$$

Cosine similarity



# Weights for internal validation

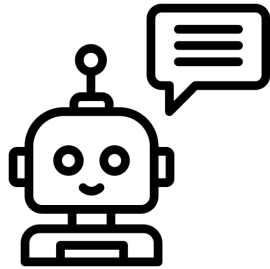
*#for quali data, subtype + geom > sum of other weights (12)*

```
{  
  'hazards' : 2,  
  'country' : 2,  
  'startYear' : 2,  
  'startMonth' : 1,  
  'startDay' : 1,  
  'endYear' : 1,  
  'endMonth' : 1,  
  'endDay' : 1,  
  'impactSubtype' : 10,  
  'impactUnit' : 1,  
  'geometry' : 5,  
  'impactValue' : 0  
}
```

*#for quanti data, 2 of impSub, impUn or impVal > sum of other weights (16)*

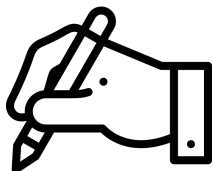
```
{  
  'hazards' : 2,  
  'country' : 2,  
  'startYear' : 2,  
  'startMonth' : 1,  
  'startDay' : 1,  
  'endYear' : 1,  
  'endMonth' : 1,  
  'endDay' : 1,  
  'impactSubtype' : 12,  
  'impactUnit' : 12,  
  'geometry' : 5,  
  'impactValue' : 6  
}
```

# Internal Validation



LLM

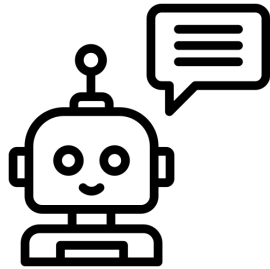
△ appealCode	△ reportDate	△ impactSubtype	# impactValue	# impactValueMin	# impactValueMax	△ impactUnit	...	📍 location	📍 country	📍 hazards
MDRSD034	2025-03-17	Affected People	460743.0	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Conflict']
MDRSD034	2025-03-17	Affected People	696000.0	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Conflict']
MDRSD034	2025-03-17	Displaced People	Missing value	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Conflict']
MDRSD034	2025-03-17	Road Infrastructure	Missing value	Missing value	Missing value	roads		📍	['Sudan']	['Flood']
MDRSD034	2025-03-17	Human Health and Wellbeing	Missing value	Missing value	Missing value	Missing value		📍	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Healthcare Infrastructure	80.0	Missing value	Missing value	% of healthcare facilities		📍	['Sudan']	['Conflict']
MDRSD034	2025-03-17	Displaced People	10000000.0	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Displaced People	10000000.0	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Other Economic and Livelihood Impa	24800000.0	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Other Economic and Livelihood Impa	1000.0	1000.0	Missing value	people		📍	['Sudan']	['Flood', 'Epidemics']



Manually  
Labelled

△ appealCode	△ reportDate	△ impactSubtype	# impactValue	# impactValueMin	# impactValueMax	△ impactUnit	...	📍 location	📍 country	📍 hazards
MDRSD034	2025-03-17	Affected People	460743.0	Missing value	Missing value	people		['Kassala', 'Gederef', 'Nile', 'Northern']	['Sudan']	['Flood', 'Epidemics', 'Conflict']
MDRSD034	2025-03-17	Injured People	1246.0	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Displaced People	178460.0	Missing value	Missing value	people		['Northern States', 'River Nile State']	['Sudan']	['Flood']
MDRSD034	2025-03-17	Homeless People	185805.0	Missing value	Missing value	people		['Red Sea', 'North Darfur', 'White Nile']	['Sudan']	['Flood', 'Epidemics', 'Conflict']
MDRSD034	2025-03-17	Human Deaths	64.0	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Epidemics', 'Conflict']
MDRSD034	2025-03-17	Infected and Ill People	57000.0	1000.0	1000.0	cases		['White Nile', 'Gazira', 'Gadarief', 'Kas']	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Residential Buildings	39081.0	Missing value	Missing value	homes		['Northern State', 'Red Sea State', 'Ri']	['Sudan']	['Flood', 'Conflict', 'Epidemics']
MDRSD034	2025-03-17	Water, Sanitation, and Hygiene Infra	1380.0	Missing value	Missing value	water, sanitation and hygiene facilities		['Port Sudan', 'Tokar', 'Aqeeq']	['Sudan']	['Flood']
MDRSD034	2025-03-17	Healthcare Infrastructure	80.0	Missing value	Missing value	% of healthcare facilities		📍	['Sudan']	['Conflict', 'Flood', 'Epidemics']
MDRSD034	2025-03-17	Road Infrastructure	Missing value	Missing value	Missing value	Missing value		['Northern State', 'Red Sea State', 'At']	['Sudan']	['Flood']

# Internal Validation

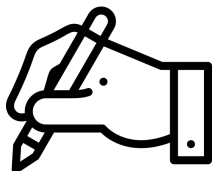


LLM

△ appealCode	△ reportDate	△ impactSubtype	# impactValue	# impactValueMin	# impactValueMax	△ impactUnit	...	📍 location	📍 country	📍 hazards
MDRSD034	2025-03-17	Affected People	460743.0	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Conflict']
MDRSD034	2025-03-17	Affected People	696000.0	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Conflict']
MDRSD034	2025-03-17	Displaced People	Missing value	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Conflict']
MDRSD034	2025-03-17	Road Infrastructure	Missing value	Missing value	Missing value	roads		📍	['Sudan']	['Flood']
MDRSD034	2025-03-17	Human Health and Wellbeing	Missing value	Missing value	Missing value	Missing value		📍	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Healthcare Infrastructure	80.0	Missing value	Missing value	% of healthcare facilities		📍	['Sudan']	['Conflict']
MDRSD034	2025-03-17	Displaced People	10000000.0	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Displaced People	10000000.0	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Other Economic and Livelihood Impa	24800000.0	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Other Economic and Livelihood Impa	1000.0	1000.0	Missing value	people		📍	['Sudan']	['Flood', 'Epidemics']



How to match same impacts ?  
Can we estimate the accuracy of the LLM ?



Manually  
Labelled

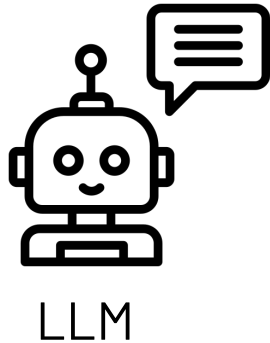
△ appealCode	△ reportDate	△ impactSubtype	# impactValue	# impactValueMin	# impactValueMax	△ impactUnit	...	📍 location	📍 country	📍 hazards
MDRSD034	2025-03-17	Affected People	460743.0	Missing value	Missing value	people		['Kassala', 'Gederef', 'Nile', 'Northern']	['Sudan']	['Flood', 'Epidemics', 'Conflict']
MDRSD034	2025-03-17	Injured People	1246.0	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Displaced People	178460.0	Missing value	Missing value	people		['Northern States', 'River Nile State']	['Sudan']	['Flood']
MDRSD034	2025-03-17	Homeless People	185805.0	Missing value	Missing value	people		['Red Sea', 'North Darfur', 'White Nile']	['Sudan']	['Flood', 'Epidemics', 'Conflict']
MDRSD034	2025-03-17	Human Deaths	64.0	Missing value	Missing value	people		📍	['Sudan']	['Flood', 'Epidemics', 'Conflict']
MDRSD034	2025-03-17	Infected and Ill People	57000.0	1000.0	1000.0	cases		['White Nile', 'Gazira', 'Gadarief', 'Kas']	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Residential Buildings	39081.0	Missing value	Missing value	homes		['Northern State', 'Red Sea State', 'Ri']	['Sudan']	['Flood', 'Conflict', 'Epidemics']
MDRSD034	2025-03-17	Water, Sanitation, and Hygiene Infra	1380.0	Missing value	Missing value	water, sanitation and hygiene facilities		['Port Sudan', 'Tokar', 'Aqeeq']	['Sudan']	['Flood']
MDRSD034	2025-03-17	Healthcare Infrastructure	80.0	Missing value	Missing value	% of healthcare facilities		📍	['Sudan']	['Conflict', 'Flood', 'Epidemics']
MDRSD034	2025-03-17	Road Infrastructure	Missing value	Missing value	Missing value	Missing value		['Northern State', 'Red Sea State', 'At']	['Sudan']	['Flood']

# Internal Validation

1. Separate Quanti/Quali
2. Vectorize
3. Compute column similarity  
(Cosine similarity, IoT for geometry, normalized difference for impactValue)
4. Total similarity = weighted sum of individual similarities

Maximize the total similarity in the row association  
Threshold of **0.5** for matching

# Internal Validation



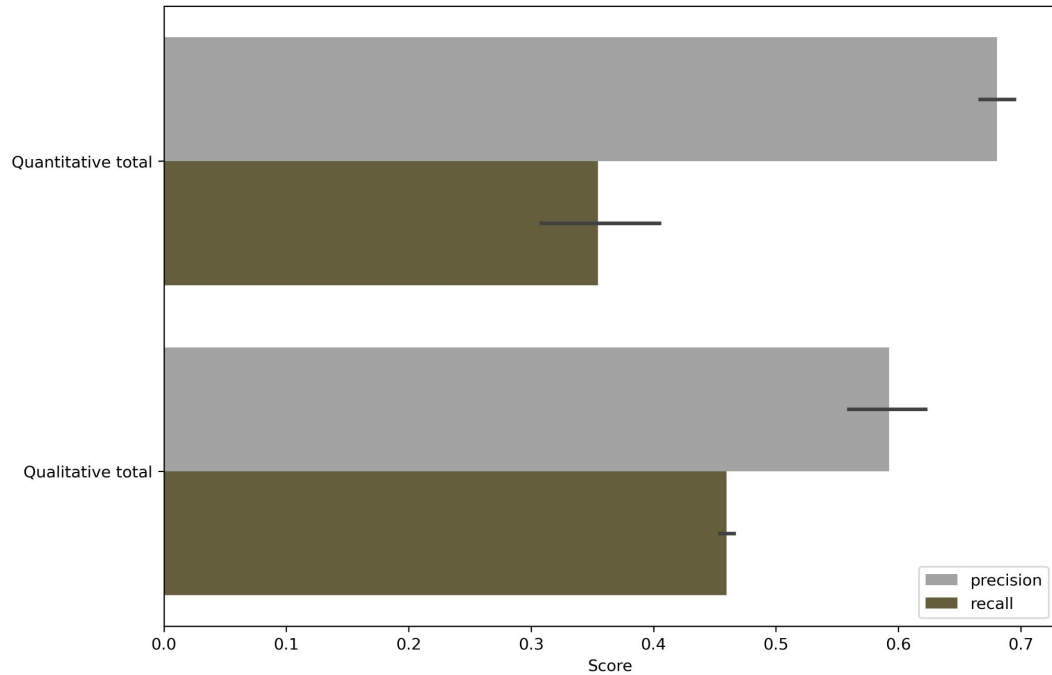
appealCode	reportDate	impactSubtype	# impactValue	# impactValueMin	# impactValueMax	impactUnit	location	country	hazards
MDRSD034	2025-03-17	Affected People	460743.0	Missing value	Missing value	people		['Sudan']	['Flood', 'Conflict']
MDRSD034	2025-03-17	Affected People	696000.0	Missing value	Missing value	people		['Sudan']	['Flood', 'Conflict']
MDRSD034	2025-03-17	Displaced People	Missing value	Missing value	Missing value	people		['Sudan']	['Flood', 'Conflict']
MDRSD034	2025-03-17	Road Infrastructure	Missing value	Missing value	Missing value	roads		['Sudan']	['Flood']
MDRSD034	2025-03-17	Human Health and Wellbeing	Missing value	Missing value	Missing value	Missing value		['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Healthcare Infrastructure	80.0	Missing value	Missing value	% of healthcare facilities		['Sudan']	['Conflict']
MDRSD034	2025-03-17	Displaced People	10000000.0	Missing value	Missing value	people		['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Displaced People	10000000.0	Missing value	Missing value	people		['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Other Economic and Livelihood Impa	24800000.0	Missing value	Missing value	people		['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Other Economic and Livelihood Impa	1000.0	1000.0	Missing value	people		['Sudan']	['Flood', 'Epidemics']



appealCode	reportDate	impactSubtype	# impactValue	# impactValueMin	# impactValueMax	impactUnit	location	country	hazards
MDRSD034	2025-03-17	Affected People	460743.0	Missing value	Missing value	people	['Kassala', 'Gederef', 'Nile', 'Northern']	['Sudan']	['Flood', 'Epidemics', 'Conflict']
MDRSD034	2025-03-17	Injured People	1240.0	Missing value	Missing value	people		['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Displaced People	178460.0	Missing value	Missing value	people	['Northern States', 'River Nile State']	['Sudan']	['Flood']
MDRSD034	2025-03-17	Homeless People	185805.0	Missing value	Missing value	people	['Red Sea', 'North Darfur', 'White Nile']	['Sudan']	['Flood', 'Epidemics', 'Conflict']
MDRSD034	2025-03-17	Human Deaths	64.0	Missing value	Missing value	people		['Sudan']	['Flood', 'Epidemics', 'Conflict']
MDRSD034	2025-03-17	Infected and Ill People	57000.0	1000.0	1000.0	cases	['White Nile', 'Gazira', 'Gadarief', 'Kas']	['Sudan']	['Flood', 'Epidemics']
MDRSD034	2025-03-17	Residential Buildings	39081.0	Missing value	Missing value	homes	['Northern State', 'Red Sea State', 'Ri']	['Sudan']	['Flood', 'Conflict', 'Epidemics']
MDRSD034	2025-03-17	Water, Sanitation, and Hygiene Infra	1380.0	Missing value	Missing value	water, sanitation and hygiene facilities	['Port Sudan', 'Tokar', 'Aqeeq']	['Sudan']	['Flood']
MDRSD034	2025-03-17	Healthcare Infrastructure	80.0	Missing value	Missing value	% of healthcare facilities		['Sudan']	['Conflict', 'Flood', 'Epidemics']
MDRSD034	2025-03-17	Road Infrastructure	Missing value	Missing value	Missing value	Missing value	['Northern State', 'Red Sea State', 'At']	['Sudan']	['Flood']

# Sensitivity analysis

Internally  
(Same model, repeated extraction)



Externally  
(Different models)

