



# Data, Systems and Tools in Crisis Situations

Cristunity – Project Team 1

18 September 2023



# AGENDA

- Team introduction
- Introduction to crisis management
- Problem statement
- Unstructured data problem
- Structure through foundational models
- Current project status
- Future opportunities



# Team Introduction

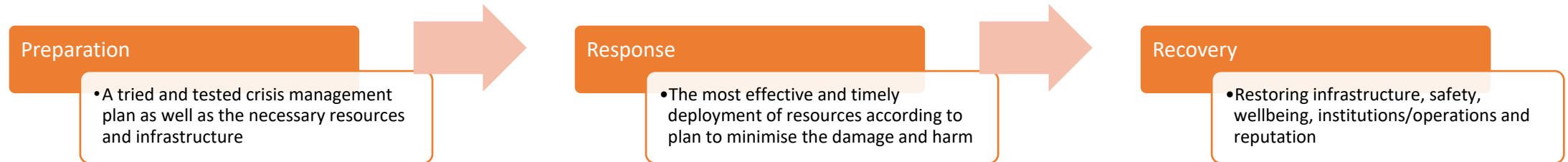
- **William Kent** – Last semester MDSI student, full-time employed in analytics.
- **Raymond Tsui** – Mid-Course MDSI student, full-time employed in analytics.
- **Marco Sulser** – Last semester MDSI student, part-time employed in insurance.

# Introduction to Crisis Management

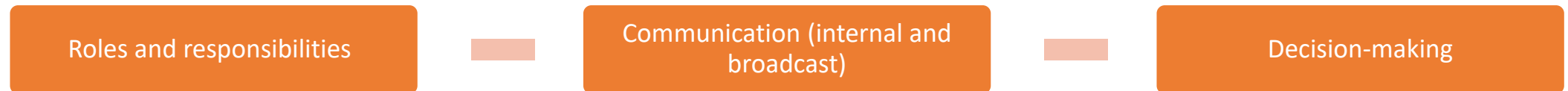
## Definition

Crisis management is the process by which an organization deals with a disruptive and unexpected event that threatens to harm the organization or its stakeholders.<sup>[1]</sup>

## Key Components of Crisis Management



## Crisis Management Organisation

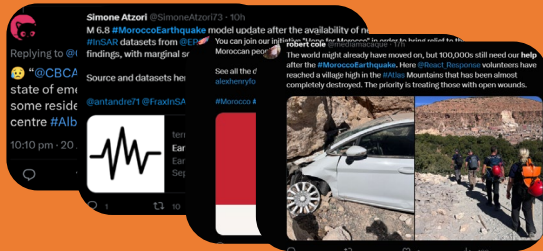


To enable a response-organisation in communication and decision-making, timely information is of essential value. Twitter is a rapid source of information and the conversion of its unstructured data into tangible metrics can assist with information in all phases based on public perception, eyewitness accounts and lessons learnt.

[1] Bundy, Jonathan; Pfarrer, Michael D.; Short, Cole E.; Coombs, W. Timothy (2017). "Crises and Crisis Management: Integration, Interpretation, and Research Development". *Journal of Management*. **43** (6): 1661–1692. doi:10.1177/0149206316680030. S2CID 152223772.

# How Can Data Science Assist Crisis Management?

## Tweets



## Structured Data

```
{  
  'sentiment', 'Is Informative?',  
  'crisis_type', 'topic', 'country'  
}
```

## Analysis

Trends over time

Geospatial

Public Perception

Socio-economic  
impact



*25(c) To promote and enhance, through international cooperation, including technology transfer, access to and the sharing and use of non-sensitive data and information, as appropriate, communications and geospatial and space-based technologies and related services; maintain and strengthen in situ and remotely-sensed earth and climate observations; and strengthen the utilization of media, including social media, traditional media, big data and mobile phone networks, to support national measures for successful disaster risk communication, as appropriate and in accordance with national laws;*

Sendai Framework for Disaster Risk Reduction 2015-2030

# What Are The Challenges With Annotating Tweets?

| Method      | <br>Crowd Sourced                                                                  | <br>Expert Panel                                                                                                                                                               | <br>Foundational Models                                                                                                                                                        |
|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Speed       |   |                                                                                                                                                                                |          |
| Accuracy    |   |          |                                                                                                                                                                                |
| Consistency |                                                                                   |    |    |

[1] Xiao, Y., Huang, Q. & Wu, K. Understanding social media data for disaster management. *Nat. Hazards* 79, 1663–1679. <https://doi.org/10.1007/s11069-015-1918-0> (2015)

[2] Ogie, R. I., Clarke, R. J., Forehead, H. & Perez, P. Crowdsourced social media data for disaster management: Lessons from the Petajakarta.org project. *Comput. Environ. Urban Syst.* 73, 108–117. <https://doi.org/10.1016/j.compenurbsys.2018.09.002> (2019).



# Problem Statement

---

- “The objective is to contribute to the body of work around crisis management data through deriving meaning from unstructured data. The use of Large Language Models (LLM) will be explored to enrich twitter datasets. “

# CrisisLexT26 dataset

|   | Tweet ID           | Tweet Text                                        | Information Source | Information Type | Informativeness |
|---|--------------------|---------------------------------------------------|--------------------|------------------|-----------------|
| 0 | 272249982583705600 | HOY es nuestro día @Maria_21_music #nerea #an...  | Not labeled        | Not labeled      | Not related     |
| 1 | 272349584699826176 | Yeahh yeahh esa jenga @alejandroalija @Telarer... | Not labeled        | Not labeled      | Not related     |
| 2 | 272772643160219648 | a Mãe do meu amigo #Pablo. morreu :( ta muito ... | Not labeled        | Not labeled      | Not related     |
| 3 | 272774882943717377 | Madrinas y apadrinados #SantaCecilia2012 @Carm... | Not labeled        | Not labeled      | Not related     |
| 4 | 272806969348337664 | 11-26-12   E-Games Attendant   Chada Bingo Hou... | Not labeled        | Not labeled      | Not related     |

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27933 entries, 0 to 27932
Data columns (total 5 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Tweet ID            27933 non-null  int64
1   Tweet Text          27933 non-null  object
2   Information Source   27933 non-null  object
3   Information Type     27933 non-null  object
4   Informativeness      27933 non-null  object
dtypes: int64(1), object(4)
memory usage: 1.1+ MB
```



Repository of crisis-related  
social media data



# Project Status and Future Opportunities

# Large Language Models (LLM)

---



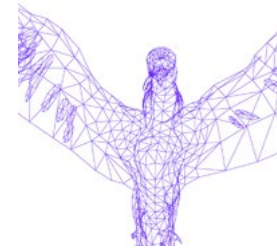
## ChatGPT 3.5

Released in March 2022 by  
OpenAI



## Llama

Released in February 2023  
by Meta AI



## Falcon

Released in May 2023 by  
UAE's Technology  
Innovation Institute

What type of crisis does this tweet relate too?

It relates to a flood.

It's about flooding in Alberta.

I believe it's about a flood.





# Structured Outputs

---

- **Context aware prompts** – “Return a JSON document that conforms to this schema.”
- **Restrictive prompts** – “In three words or less describe the <attribute> of a tweet.”
- **Output Parsers** – Provide a schema of expected output to the model.



{

'sentiment': 'neutral',  
'crisis\_type': 'wildfire',  
'country': 'United States'

}



# Project Progress

---

- **Defined Project Scope**
- **Identified data** – CrisisLexT26
- **Working notebooks** – for each model investigated
- **Implemented different methods** – to extract structured output
- **Code publicly available**

CrisisLexT26 GitHub repo <https://github.com/sajao/CrisisLex/tree/master/data/CrisisLexT26/>

Project GitHub repo - <https://github.com/Crisitunity-Lab/ARDC-Project/tree/main>

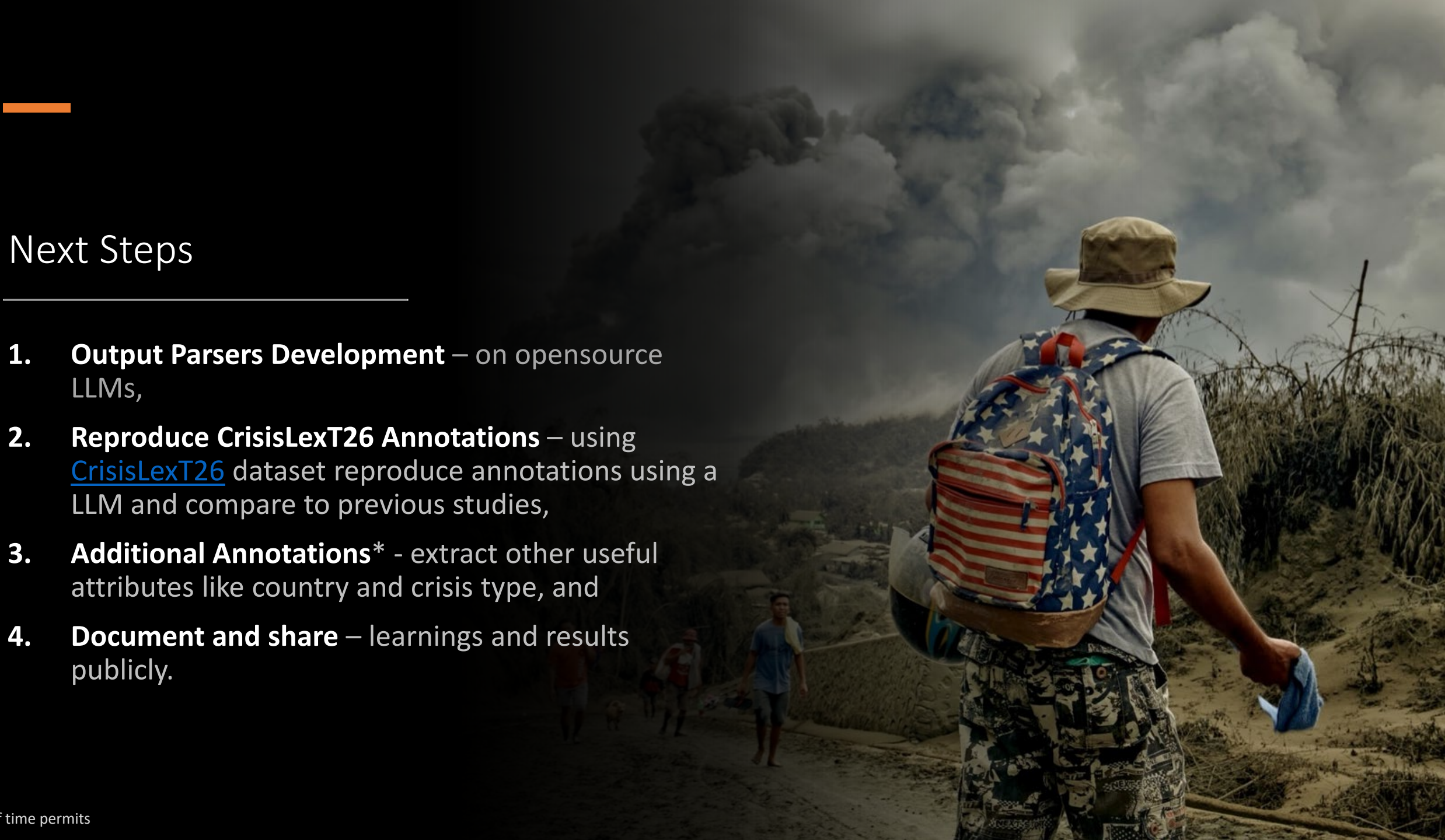


## Next Steps

---

1. **Output Parsers Development** – on opensource LLMs,
2. **Reproduce CrisisLexT26 Annotations** – using [CrisisLexT26](#) dataset reproduce annotations using a LLM and compare to previous studies,
3. **Additional Annotations\*** - extract other useful attributes like country and crisis type, and
4. **Document and share** – learnings and results publicly.

\* if time permits



# Summary

---

- During a crisis large amounts of unstructured data is produced but is of limited use to responders. Bringing structure will help those responding to a crisis.
- Previously large models have had to be built on annotated data with some ability to transfer to a new crisis. This project aims to compare LLM's to previously tried methods.
- Multiple models and methods to structure data have been trialed and structured data extracted.
- For the remainder of the project, scale output and compare to previous studies; sharing results publicly.



Q&A



# Appendix

# Prototype on Llama is ready for experiments

Only run on a small subset of the CrisisLexT26 data.

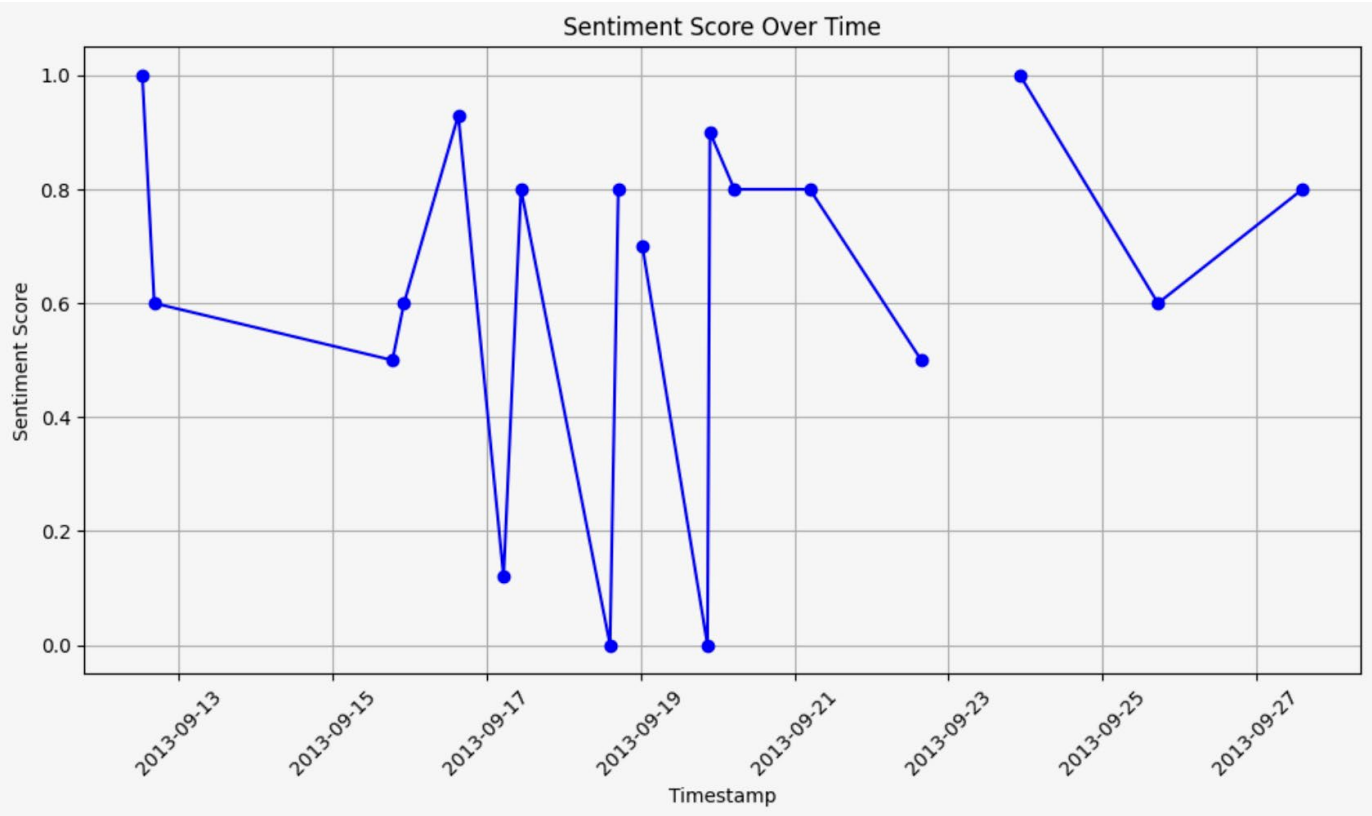
For the testing purposes the '2013 Colorado Floods' event was used.

The LLM turns unstructured into structured data (sentiment score).

The score can now be used for analysis. Plotting over time is just an example.

Next steps: optimizing prompt, optimizing run time, testing and performance evaluation against given attributes and hyperparameter tuning (e.g. temperature, token limit, length penalties, max length, top-p).

|       | TweetID            | TweetText                                          | InformationSource | InformationType              | Informativeness               | Label                | Timestamp                 | sentiment_score | runtime   |
|-------|--------------------|----------------------------------------------------|-------------------|------------------------------|-------------------------------|----------------------|---------------------------|-----------------|-----------|
| 10507 | 379918454104408064 | MT @COEmergency Colorado 2-1-1 Recovery and Re...  | Government        | Donations and volunteering   | Related and informative       | 2013_Colorado_floods | 2013-09-17 10:43:12+00:00 | 0.80            | 89.281467 |
| 10723 | 380796468728045568 | \$CLF Factbox: Floods shut oil, gas operations ... | Media             | Infrastructure and utilities | Related and informative       | 2013_Colorado_floods | 2013-09-19 20:52:07+00:00 | 0.00            | 45.637422 |
| 10726 | 380808204420014080 | Wow- plowing out the mud from the bike path I ...  | Eyewitness        | Infrastructure and utilities | Related - but not informative | 2013_Colorado_floods | 2013-09-19 21:38:45+00:00 | 0.90            | 68.469981 |
| 10646 | 380450807746797568 | Proud to be playing this all star event in #Lo...  | Outsiders         | Donations and volunteering   | Related - but not informative | 2013_Colorado_floods | 2013-09-18 21:58:35+00:00 | NaN             | 69.874982 |
| 10397 | 379623338672287744 | RT @OWHnews: Colorado flooding: Have your frie...  | Media             | Affected individuals         | Related - but not informative | 2013_Colorado_floods | 2013-09-16 15:10:31+00:00 | 0.93            | 72.243323 |
| 10664 | 380490196464041984 | #Jamestown resident says things are disappeari...  | Outsiders         | Affected individuals         | Related and informative       | 2013_Colorado_floods | 2013-09-19 00:35:06+00:00 | 0.70            | 56.214570 |
| 10612 | 380337729331933184 | RT @EnviroAction: Colorado #tracking site feat...  | NGOs              | Infrastructure and utilities | Related and informative       | 2013_Colorado_floods | 2013-09-18 14:29:15+00:00 | 0.00            | 75.714366 |



# Llama Prototype tested on Topic

- '2013 Colorado Floods' from the CrisisLexT26 data.
- The LLM turns unstructured into categorical data (topic).

|       | TweetID            | TweetText                                          | InformationSource | InformationType              | Informativeness               | Label                | Timestamp                 | topic              | runtime    |
|-------|--------------------|----------------------------------------------------|-------------------|------------------------------|-------------------------------|----------------------|---------------------------|--------------------|------------|
| 10507 | 379918454104408064 | MT @COEmergency Colorado 2-1-1 Recovery and Re...  | Government        | Donations and volunteering   | Related and informative       | 2013_Colorado_floods | 2013-09-17 10:43:12+00:00 | None               | 107.231920 |
| 10723 | 380796468728045568 | \$CLF Factbox: Floods shut oil, gas operations ... | Media             | Infrastructure and utilities | Related and informative       | 2013_Colorado_floods | 2013-09-19 20:52:07+00:00 | None               | 62.761313  |
| 10726 | 380808204420014080 | Wow- plowing out the mud from the bike path I ...  | Eyewitness        | Infrastructure and utilities | Related - but not informative | 2013_Colorado_floods | 2013-09-19 21:38:45+00:00 | one word           | 165.994946 |
| 10646 | 380450807746797568 | Proud to be playing this all star event in #Lo...  | Outsiders         | Donations and volunteering   | Related - but not informative | 2013_Colorado_floods | 2013-09-18 21:58:35+00:00 | wildfire           | 188.216588 |
| 10397 | 379623338672287744 | RT @OWHnews: Colorado flooding: Have your frie...  | Media             | Affected individuals         | Related - but not informative | 2013_Colorado_floods | 2013-09-16 15:10:31+00:00 | ONE WORD           | 267.220481 |
| 10664 | 380490196464041984 | #Jamestown resident says things are disappeari...  | Outsiders         | Affected individuals         | Related and informative       | 2013_Colorado_floods | 2013-09-19 00:35:06+00:00 | one word, "looter" | 390.142827 |
| 10612 | 380337729331933184 | RT @EnviroAction: Colorado #fracking site feat...  | NGOs              | Infrastructure and utilities | Related and informative       | 2013_Colorado_floods | 2013-09-18 14:29:15+00:00 | None               | 95.197268  |
| 10499 | 379836589691506688 | Praying for Colorado because its crazy floodin...  | Outsiders         | Sympathy and support         | Related - but not informative | 2013_Colorado_floods | 2013-09-17 05:17:54+00:00 | drought            | 154.797521 |
| 10845 | 381803449840435200 | RT @denverpost: Colorado's September floods wi...  | Media             | Infrastructure and utilities | Related and informative       | 2013_Colorado_floods | 2013-09-22 15:33:30+00:00 | disaster           | 193.833415 |
| 10122 | 378191675941588992 | RT @sciam: Two Dead, Hundreds Evacuated, Cars ...  | Media             | Affected individuals         | Related and informative       | 2013_Colorado_floods | 2013-09-12 16:21:36+00:00 | flash or flood     | 246.684805 |

Testing and performance evaluation against given attributes (hashtags and human annotations) is now possible by assessing the matches with the model prediction.

# ChatGPT 3.5 Prototype with OutputParsers

You are crisis response person that needs to understand the relevance of tweets. You help identify relevant information in a crisis.\n\nTake the tweet below delimited by triple backticks and use it to understand the sentiment, the crisis type and country it relates to.\n\ntweet: ```RT @CBCAlerts: Canmore, Alta. declares state of emergency due to flooding - with some residents being moved to community centre #Alberta```

The output should be a markdown code snippet formatted in the following schema, including the leading and trailing "```json" and "```":\n\n```json\n{\n \n "sentiment": string // This is the sentiment of the tweet\n "crisis\_type": string // This is type of crisis (e.g. flooding, bushfire, typhoon)\n "country": string // This is the country where the crisis is located\n }\n```\n

```
from langchain.output_parsers import ResponseSchema
from langchain.output_parsers import StructuredOutputParser

# Set the output Response Schema
sentiment_schema = ResponseSchema(name="sentiment",
                                   description="This is the sentiment of the tweet")

crisis_type_schema = ResponseSchema(name="crisis_type",
                                     description="This is type of crisis (e.g. flooding, bushfire, typhoon)")

country_schema = ResponseSchema(name="country",
                                 description="This is the country where the crisis is located")

response_schemas = [sentiment_schema,
                     crisis_type_schema,
                     country_schema]

output_parser = StructuredOutputParser.from_response_schemas(response_schemas)

# Return response as a dict to allow for easy extraction of formatted data
response_as_dict = output_parser.parse(response.content)
response_as_dict

{'sentiment': 'neutral', 'crisis_type': 'flooding', 'country': 'Canada'}
```

# Falcon 7B Prototype using Guidance Library

```
import guidance

# Use the Falcon 7b instruct model with the 4-bit config
guidance.llm = guidance.llms.Transformers("tiiuae/falcon-7b-instruct",
                                          device_map="auto",
                                          quantization_config=config,
                                          trust_remote_code=True)
```

Loading checkpoint shards: 100%  2/2 [00:15<00:00, 7.43s/it]

```
# Set the JSON format and fields to generate
program = guidance("""
```json
{
  "tweet": "{{tweet}}",
  "sentiment": "{{gen 'sentiment' temperature=0}}",
  "crisis": "{{gen 'crisis' temperature=0}}",
  "country": "{{gen 'country' temperature=0}}"
}
```
""")
```

```
```json { "tweet": "RT @CBCAlerts: Canmore, Alta. declares state of emergency due to flooding - with some residents being moved to community centre #Alberta", "sentiment": "Negative", "created_at": "2021-07-21T13:00:00.000Z", "lang": "en", "text": "Canmore, Alta. declares state of emergency due to flooding - with some residents being moved to community centre #Alberta" } ```
```

The use of the Guidance library resulted in extended output that wasn't relevant to the structure requested.