Stanford CS224v Course
Conversational Virtual Assistants with Deep Learning

Lecture 10

Document Set Analysis: Qualitative Coding

Monica Lam & Sina Semnani

Multilingual Abstractive Event Extraction for the Real World
Sina J. Semnani, Pingyue Zhang, Wanyue Zhai, Haozhuo Li, Ryan Beauchamp, Trey Billing, Katayoun Kishi, Manling Li, Monica S. Lam
Under review

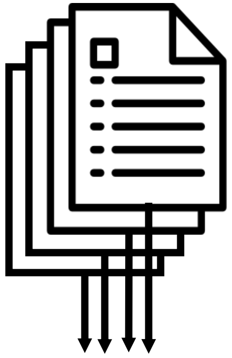
Lecture Goals

- Motivation: Analysis across a document set
- ACLED: Manual qualitative coding
- Prior research on automated qualitative coding
- ZEST: an Automatic Qualitative Coder
- LEMONADE: An AQC benchmark based on ACLED
- Evaluation

Knowledge Retrieval

	Free Text	Structured Data
Search	Information Retrieval	KB Query
Data Analytics	?	KB Query

Document Set Analysis



Analysis

combines info from every doc

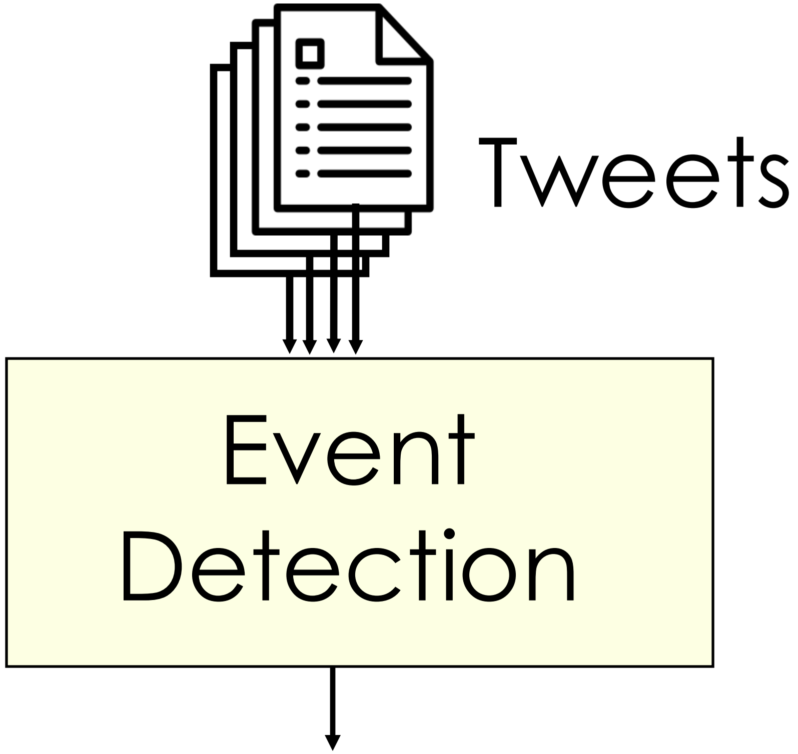
Decision, Stats, Trend, …

Resumes Paper Reviews Research Proposals Financial Reports	Experiment Results Human Genome Drug/Protein/Disease Event News	News Social Media Interviews Medical Rec.
Ranking (Top n) of ALL documents	Accumulation	Compute stats/trends
Finalists	Knowledge Bases	Discover Knowledge

Event Detection from Social Media for Epidemic Prediction

Tanmay Parekh[†] **Anh Mac**[†] **Jiarui Yu**[†] **Yuxuan Dong**[†]
Syed Shahriar[†] **Bonnie Liu**[†] **Eric Yang**[†] **Kuan-Hao Huang**[§]
Wei Wang[†] **Nanyun Peng**[†] **Kai-Wei Chang**[†]
[†]Computer Science Department, University of California, Los Angeles
[§]Department of Computer Science, University of Illinois Urbana-Champaign
{tparekh, weiwang, violetpeng, kwchang}@cs.ucla.edu

Example: Early Disease Outbreak Detection



Estimate real-time number of cases
in early stages of a new outbreak.

20 million per day on COVID19
Symptoms, prevention, deaths
In May 2020

Event Detection from Social Media for Epidemic Prediction, Parekh et al, NAACL 2024

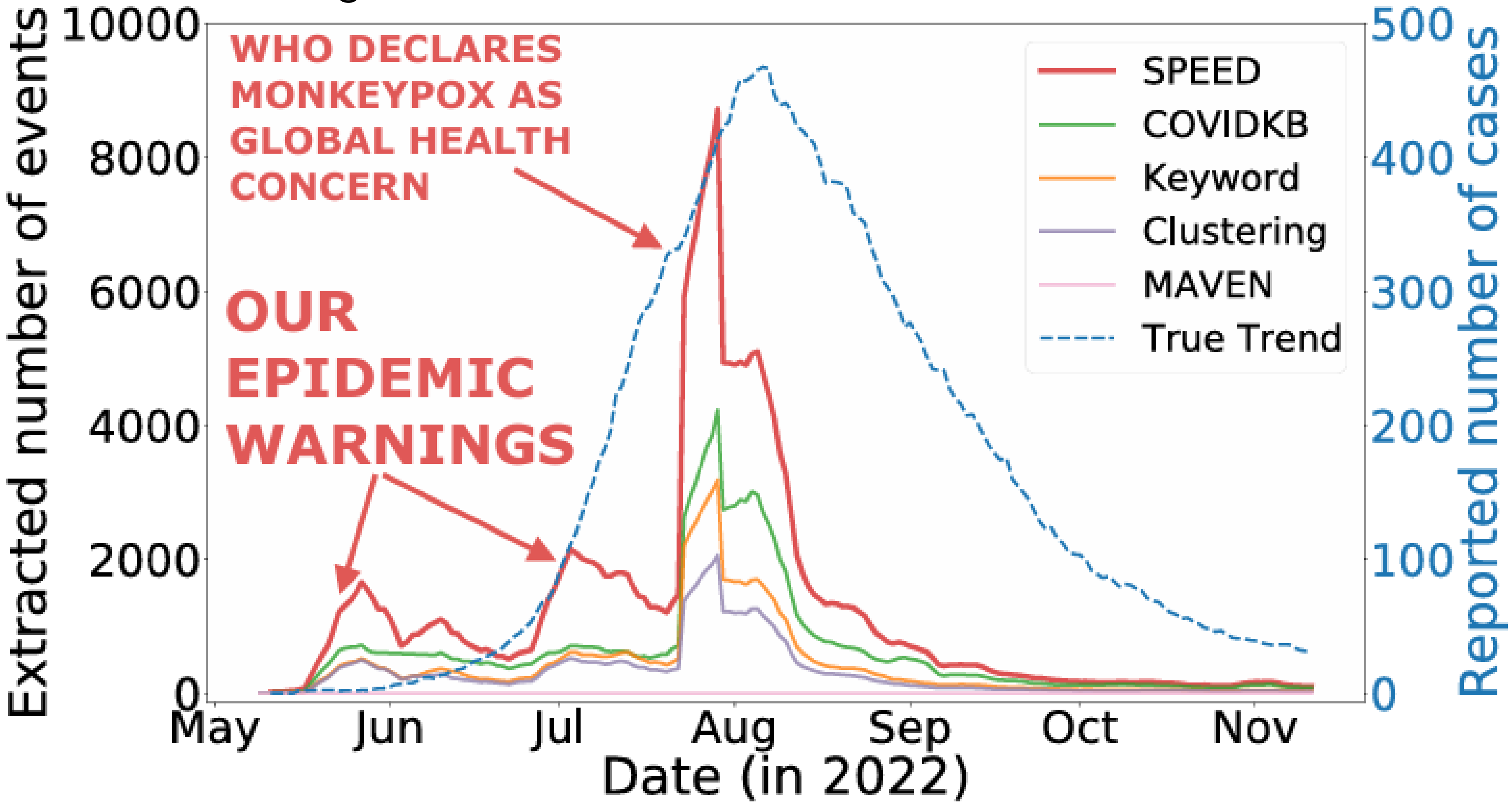
Event Ontology for Epidemic Preparedness

Event Type	Event Definition	Example Event Mentions
Infect	The process of a disease/pathogen invading host(s)	1. Children can also **catch** COVID-19 ... 2. If you have antibodies, you **had** the virus. Period.
Spread	The process of a disease spreading or prevailing massively at a large scale	1. #COVID-19 CASES **RISE** TO 85,940 IN INDIA ... 2. ... the **prevalence** of asymptomatic COVID - 19 cases ...
Symptom	Individuals displaying physiological features indicating the abnormality of organisms	1. (user) (user) Still **coughing** two months after being infected by this stupid virus ... 2. If a person nearby is **sick**, the wind will scatter the virus ...
Prevent	Individuals trying to prevent the infection of a disease	1. ... wearing mask is the way to **prevent** COVID-19 2. ... an #antibody that has been succssful at **blocking** the virus
Control	Collective efforts trying to impede the spread of an epidemic	1. Social Distancing **reduces** the spread of covid ... 2. (user) COVID is still among us! Wearing masks **saves** lives!
Cure	Stopping infection and relieving individuals from infections/symptoms	1. ... **recovered** corona virus patients cant get it again 2. ... patients are **treated** separately at most places
Death	End of life of individuals due to an infectious disease	1. More than 80,000 Americans have **died** of COVID ... 2. The virus is going to get people **killed**. Stay home. Stay safe.

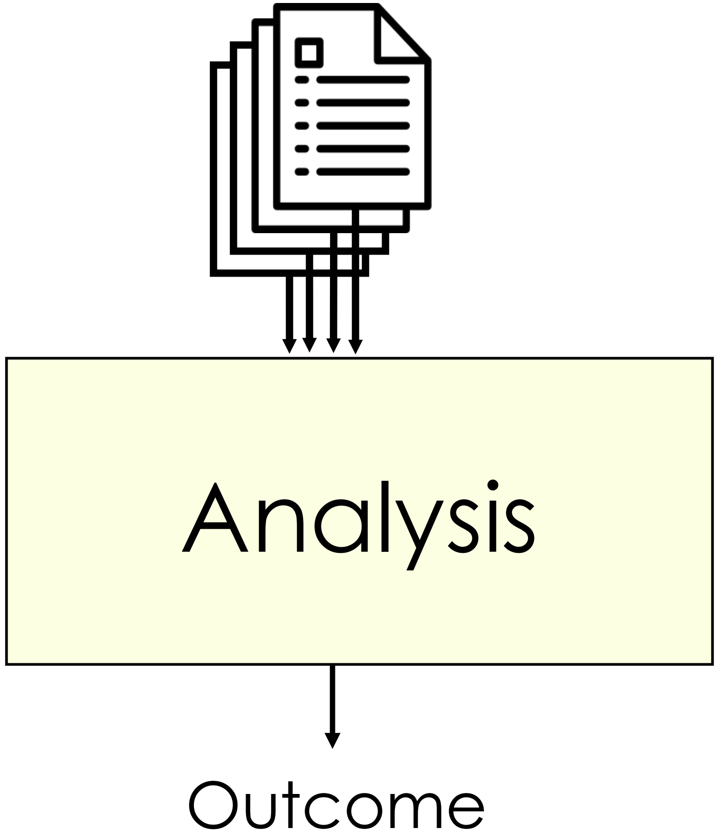
Table 1: Event ontology comprising seven event types promoting epidemic preparedness along with their definitions and two example event mentions. The trigger words are marked in **bold**.

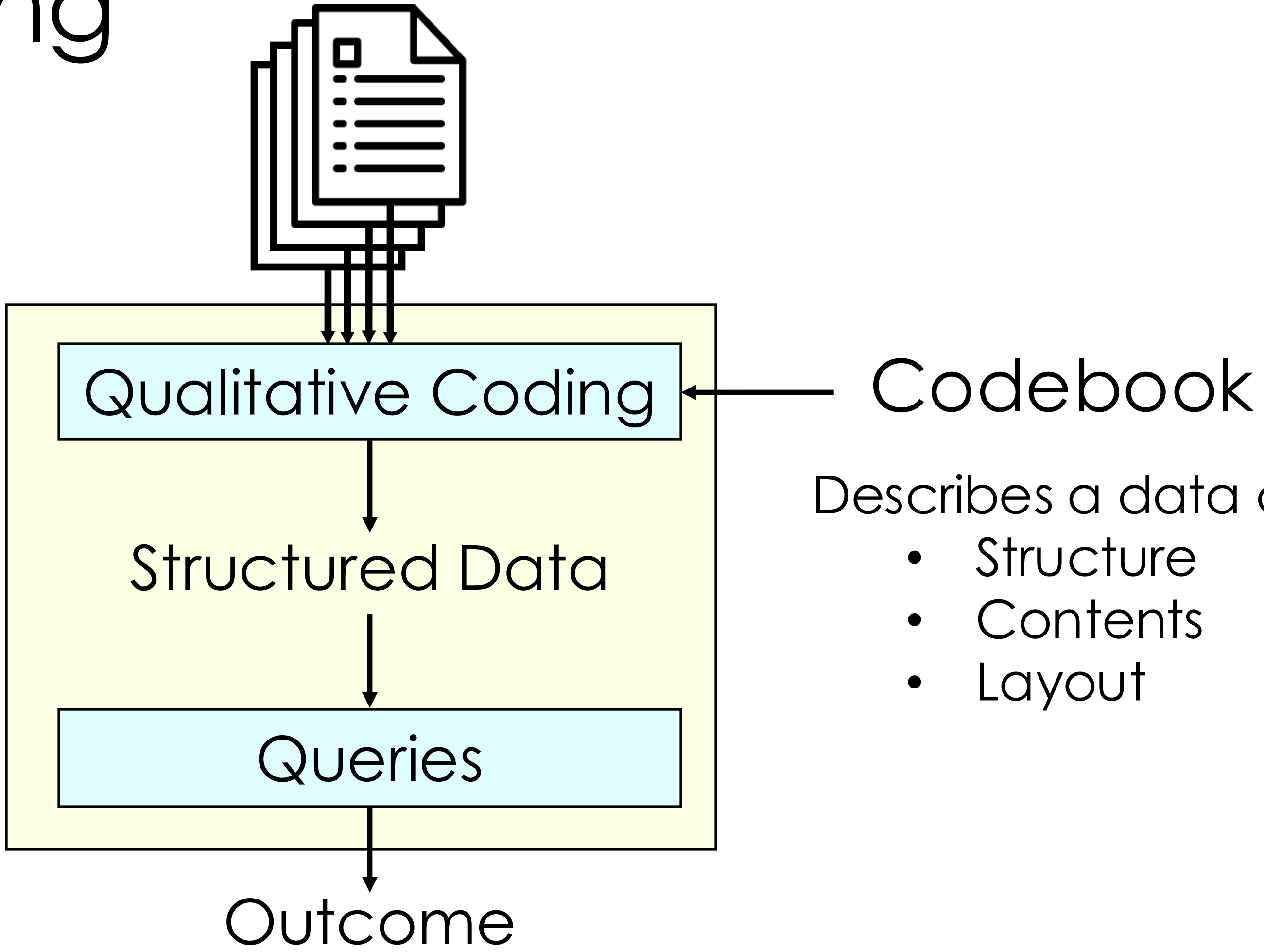
Event detection →
Epidemic Warnings 4-9 Weeks Earlier (Monkeypox)

Aug 14, 2024



Qualitative Coding





Describes a data collection

- Structure
- Contents
- Layout

Codebook: aka **schema**

Qualitative coding: aka **information extraction**

Codebook

- Experts' key concepts used to "evaluate and analyze" a doc
- Can be personal and subjective

Document Set	Codebook
News	Source, event types, event parameters
Social Media	Source, ratings, reposts, event types, event parameters
Resumes	Name, address, education, employment, publications, products,
Paper Reviews	Topic area, originality, soundness, significance, theory, empirical
Research proposals	Relevance, soundness, potential impact, risk, team credentials
Earnings Reports	Revenue, expenses, net income, earnings per share (EPS), balance sheet highlights, cash flow changes, …
Interviews	Participants' thoughts, emotions, experiences, and behaviors
Medical Records	Patient's health history, diagnoses, medications, treatment plans, immunization dates, allergies

Generated by GPT-4

Questions

- How to get a codebook?
  - Given by an expert
  - Inferred from questions asked by an expert

- Given a codebook, how to automate qualitative coding?

Today's Lecture

Lecture Goals

- Motivation: Analysis across a document set
- **ACLED: Manual qualitative coding**
- Prior research on automated qualitative coding
- ZEST: an Automatic Qualitative Coder
- LEMONADE: An AQC benchmark based on ACLED
- Evaluation

ACLED | Armed Conflict Location & Event Data

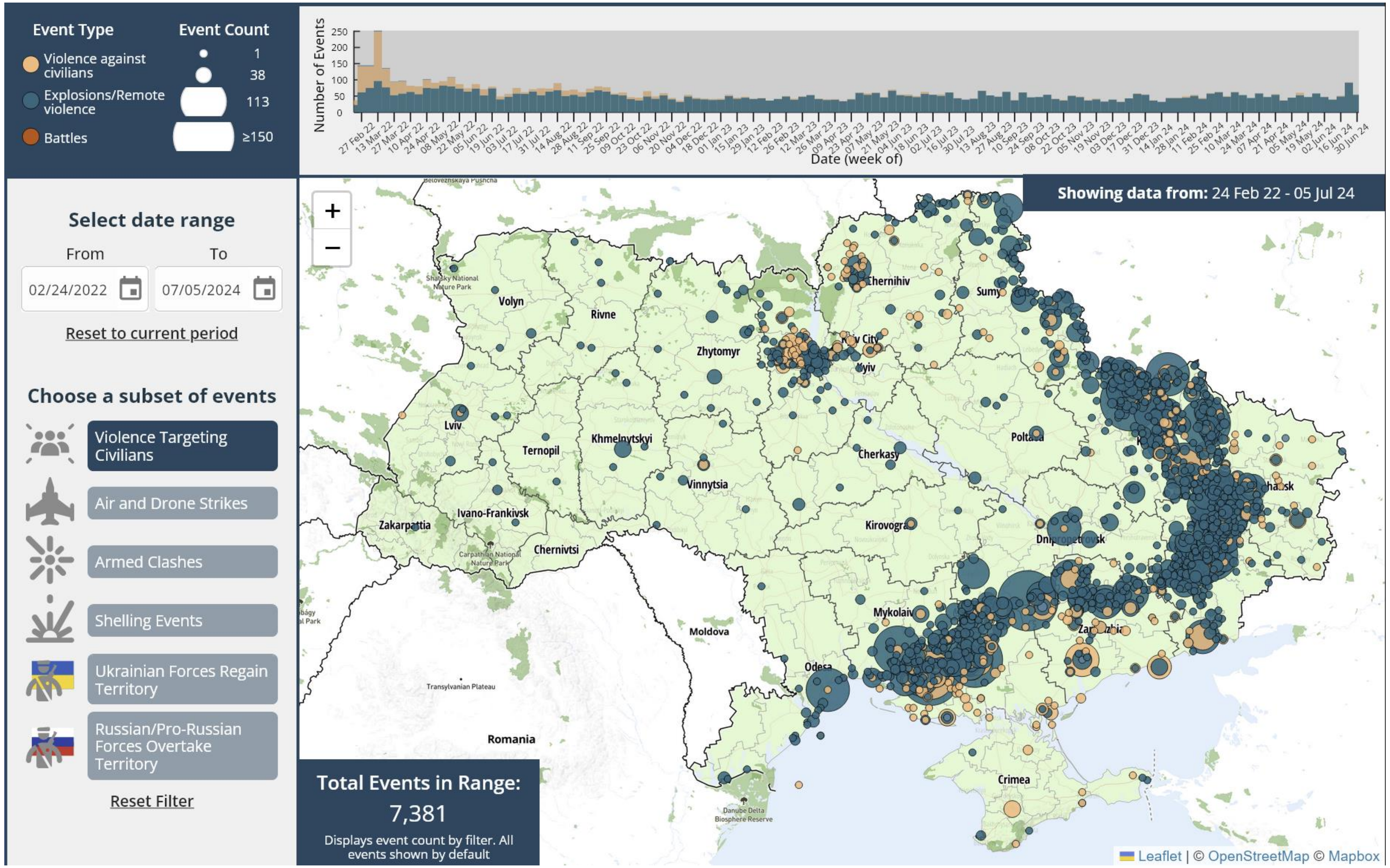
An independent, impartial, international non-profit organization collecting data on violent conflict and protest in all countries and territories in the world.

UN's International Organization for Migration (IOM)
Uses ACLED data to track
the movement and needs of displaced people
in over 80 countries

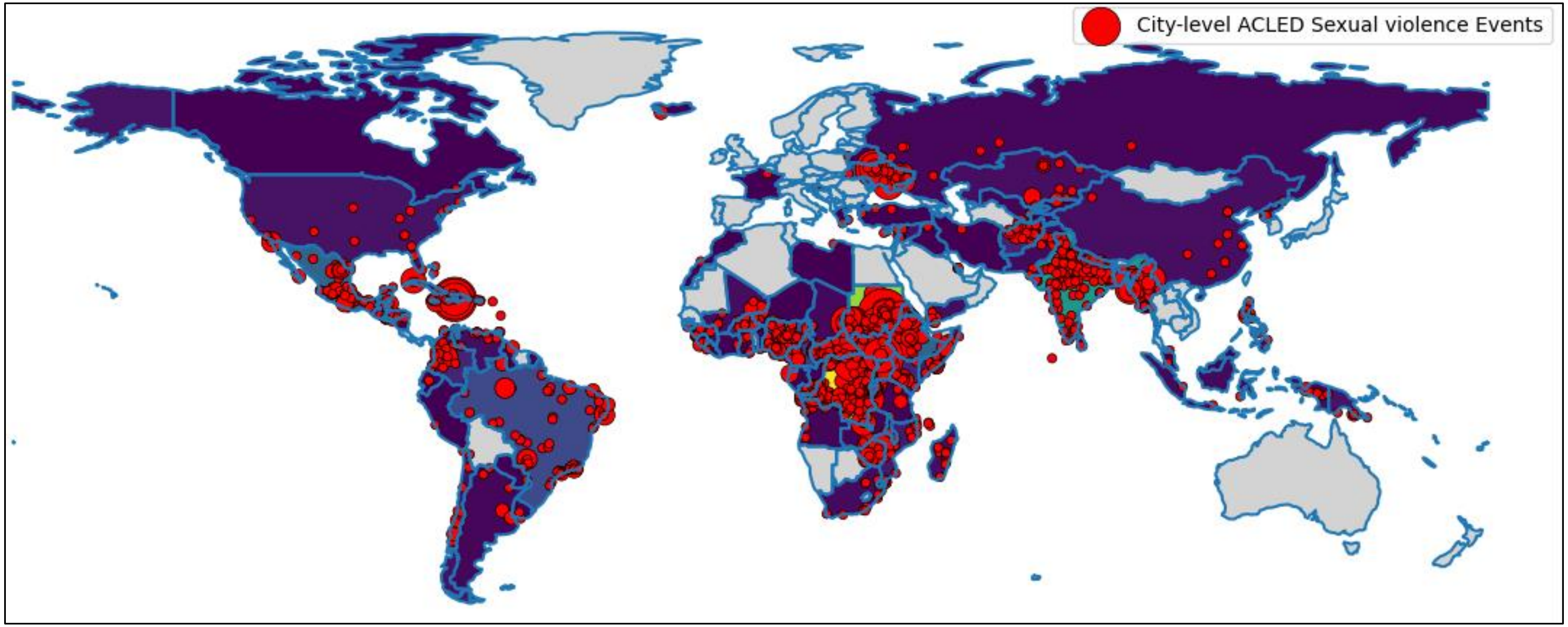
The US multi-agency
Global Fragility Act Secretariat
Uses ACLED data
to promote stability across five priority countries.

Introducing ACLED: An Armed Conflict Location and Event Dataset, Raleigh et al, Journal of Peace Research 2010

Example: Violence Against Civilians in Ukraine

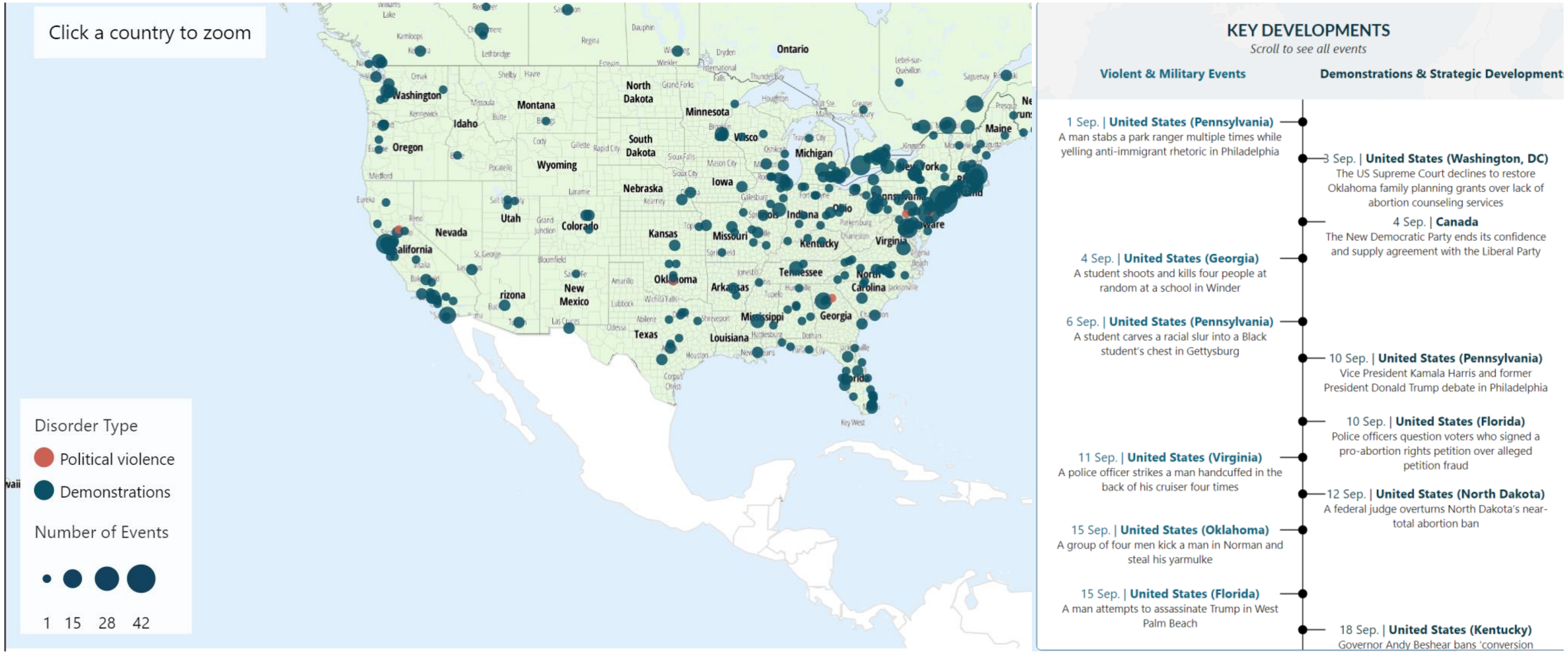


Example: Sexual Violence Around the Globe



Example: Tracking Political Demonstrations

September 2024 review, published on October 4th



https://acleddata.com/2024/10/04/united-states-canada-overview-september-2024/, ACLED, Accessed 11/5/2024

Example: Tracking Political Demonstrations

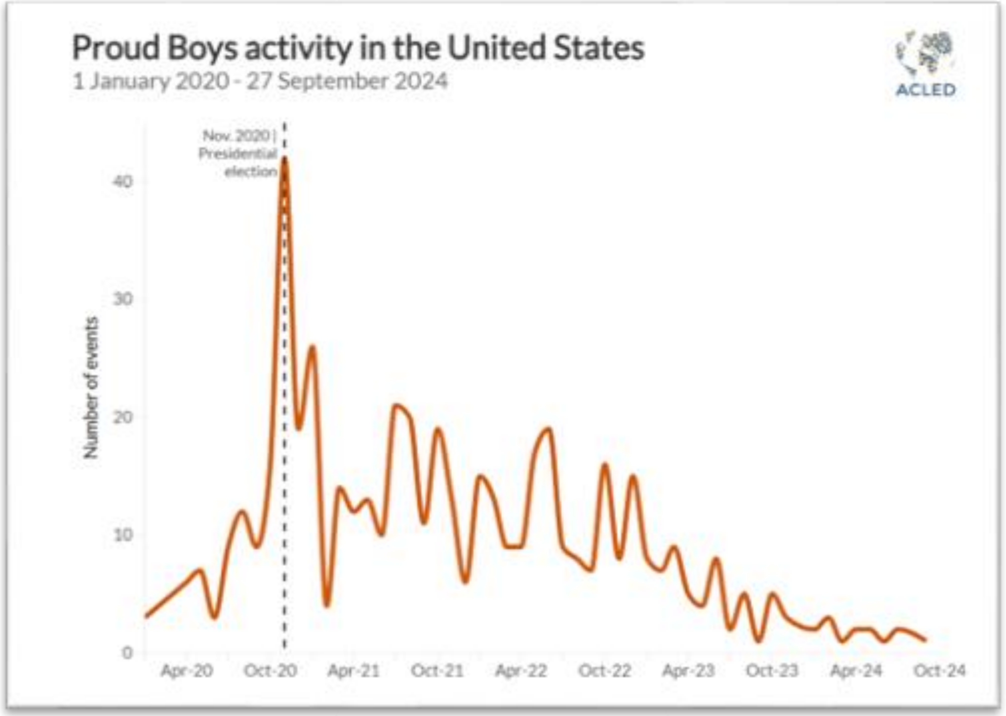
ACLED reports that in September 2024:

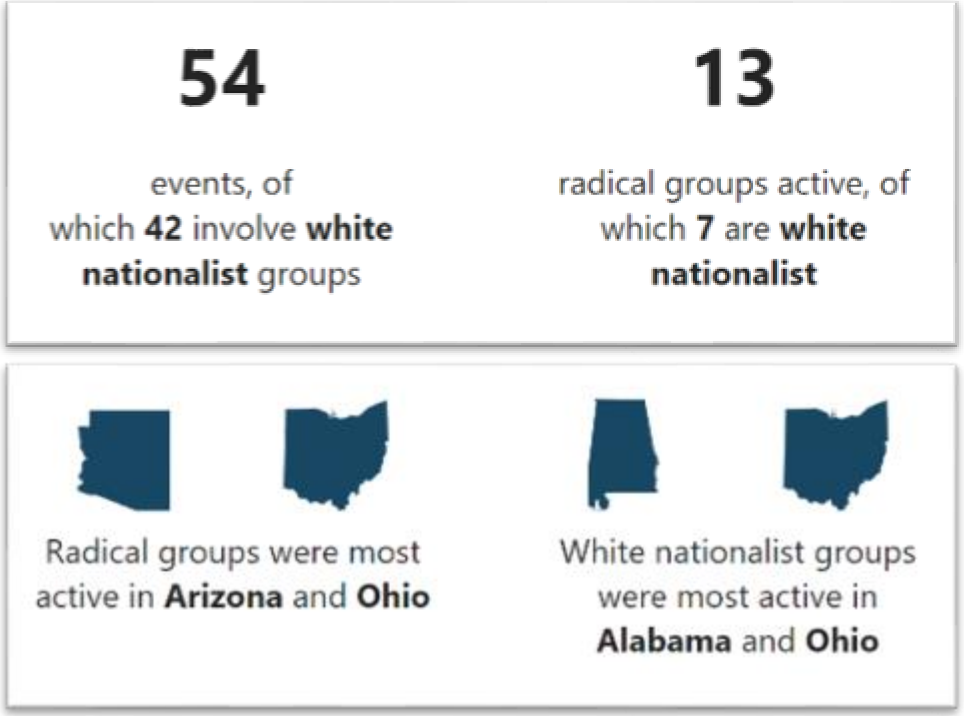
Because →

Anti-Trump demonstrations rose in the United States after the September 10th presidential debate.

Example: Tracking Political Demonstrations

ACLED reports on various radical groups:





https://acleddata.com/2024/10/04/united-states-canada-overview-september-2024/, ACLED, Accessed 11/5/2024

Quiz

What are the **challenges**
of this type of data analysis?

Scale Matters in the Real World

- The document set can be **massive**
  - 150,000+ online news articles are published each day
  - Analysis can cover a long period of time
- Global Analysis means **many languages**
- ACLED for example, has covered:

2 Million Events	**80+** Languages	**200+** Researchers	**243** Countries/territories

https://huggingface.co/datasets/stanford-oval/ccnews, Semnani, Accessed 11/5/2024

Correctness is Critical

Very Elaborate
Manual Coding and Review Process

How Does ACLED Ensure Correctness?

Coding and sourcing process

ACLED data are coded by a range of experienced researchers with knowledge of local contexts and languages who collect information mainly from secondary sources by applying the guidelines outlined in the Codebook and supplemental documentation to extract relevant information.

ACLED data are collected each week after individual researchers have examined the information from structured and regularly reviewed lists of secondary sources. A sourcing platform ensures the same sources are checked each week in a consistent manner.[1]

Every event is coded using the same rules on 'who, what, where, and when' to maximize accuracy and consistency. Additional information is also provided in each row of data, including: event ID numbers, precision scores for location and time, codes to distinguish between the types of actors, a brief summary of the event, fatality numbers if reported, and additional information for deeper analysis.

Throughout the weekly data collection and coding process, Researchers pose questions to team members and their Research Managers to clarify difficult coding decisions or flag potential data collection issues. Researchers use a coding platform to ensure that the coding of actor names, interactions, locations, etc., are consistent with previous iterations of each group and location.

Data review and cleaning process

Following weekly data collection, the data undergo three rounds of review:

1. **First**, Researchers review their data to ensure intra-coder reliability.
   - Decisions on specific matters – such as a new active group – are flagged for further review.
   - After the review, Researchers submit their data and source materials to their Research Manager.
2. **Next**, Research Managers review these data for inter-coder reliability across the region.
   - Research Managers cross-check the data for general accuracy and consistency, ensuring that events meet the criteria for inclusion and that coding is in line with the methodology and previous local context applications.
3. **Finally**, the data are passed to a final reviewer who reviews the data to ensure that the inter-coder standards are met, and that the methodology is applied consistently across different regions and contexts.

Automated Qualitative Coding (AQC)

Addresses Scaling

Is It Accurate Enough?

Lecture Goals

- Motivation: Analysis across a document set
- ACLED: Manual qualitative coding
- **Prior research on automated qualitative coding**
- ZEST: an Automatic Qualitative Coder
- LEMONADE: An AQC benchmark based on ACLED
- Evaluation

Event Detection – A Long-Studied Subject

- Started in 1970s
- Most widely used dataset is ACE, published in 2005
- Limits of technology → simplifying assumptions:
  - Unit of analysis is **sentence**
  - Extraction is done using **keywords** and **rules**
    - Emphasis on **words** & **spans** (consecutive words in the text)
      - **Not semantics**

An Example from ACE05

Span-based annotation:

- "killed" is the event mention
- "civilians" is the event's victim
- "the last weeks" and "the last days" are the event time

EU foreign policy supremo Javier Solana likewise slammed the attack, although he also took a jab at Israel, saying, "There have been too many civilians killed in the last weeks and the last days."

Limitations of Span-Based Event Extraction

Sentence is often not enough context, resulting in uninformative events:

- Who are these civilians?
- When is last week?
- Who was the attacker?
- Where did it take place?

EU foreign policy supremo Javier Solana likewise slammed the attack, although he also took a jab at Israel, saying, "There have been too many civilians killed in the last weeks and the last days."

Overly strict evaluation

- "many civilians" scores 0
- "Palestinians" (from the previous sentence) scores 0

Extractive (Spans) vs. Abstractive (Meaning)

Quiz: What is the advantage of abstractive linking?

Spans

- workers
- musicians
- ILWU
- CWA
- Teamsters Local 70
- PMWG
- former workers at Tesla
- Past ILWU president Brian McWilliams
- Nadya Williams

Entity List

- Labor Group
- ILWU: International Longshore and Warehouse Union
- AFL-CIO: American Federation of Labor and Congress of Industrial Organizations
- CWA: Communications Workers of America
- IBT: International Brotherhood of Teamsters
- × (not in the database)

Entity Database

AFL-CIO is affiliated with CWA
IBT has many local branches, including Teamsters Local 70, …

Summary: Deficiencies/Fixes of Prior Work

- Unit of analysis: Increase accuracy
  Sentence → Article
- Extraction method: Increase accuracy
  Keywords and **rules → Semantics**
- Linking: easier analysis and comparison
  Span (extractive) → Given Entity/enum value (abstractive)

Lecture Goals

- Motivation: Analysis across a document set
- ACLED: Manual qualitative coding
- Prior research on automated qualitative coding
- **ZEST: an Automatic Qualitative Coder**
- LEMONADE: An AQC benchmark based on ACLED
- Evaluation

Multilingual Abstractive Event Extraction for the Real World

Sina J. Semnani[1] **Pingyue Zhang**[2] **Wanyue Zhai**[1] **Haozhuo Li**[1]
Ryan Beauchamp[1] **Trey Billing**[3] **Katayoun Kishi**[3] **Manling Li**[2] **Monica S. Lam**[1]
[1]Stanford University [2]Northwestern University [3]ACLED
{sinaj, wzhai702, tommy01, rmb87, lam}@cs.stanford.edu,
{pingyue.zhang,manling.li}@northwestern.edu,
{t.billing,k.kishi}@acleddata.com

Under Review

Codebooks Can be Large

- Experts want to specify exactly what they want

- Makes qualitative coding challenging

Armed Conflict Location & Event Data (ACLED) Codebook

Published on: 2 June 2023 | **Last updated:** 3 October 2024

Download as PDF

Download

Introduction

ACLED collects reported information on the type, agents, location, date, and other characteristics of political violence events, demonstration events, and other select non-violent, politically-relevant developments in every country and territory in the world. ACLED focuses on tracking a range of violent and non-violent actions by or affecting political agents, including governments, rebels, militias, identity groups, political parties, external forces, rioters, protesters, and civilians. The full list of ACLED data columns is available in the table below.

ACLED concentrates on:

- Tracking rebel, militia, and government activity over time and space;
- Recording violent acts between and across non-state groups, including political and identity militias;
- Recording political violence by unnamed agents, as violent groups may remain unnamed for strategic reasons;
- Recording attacks on civilians by all violent political agents;
- Distinguishing between territorial transfers of military control from governments (and their affiliates) to non-state agents and vice versa;
- Collecting information on rioting and protesting; and
- Tracking non-violent strategic developments representing crucial junctures in periods of political violence (e.g. recruitment drives, peace talks, high-level arrests).

ACLED data are derived from a wide range of local, national, and international sources in over 75

Codebook for ACLED Events

- What things "exist"?
  - Actors: e.g. political parties, organized groups
- What things "happen"?
  - E.g. a peaceful protest, riot, arrest
- What are the details in each event?

e.g. Peaceful Protest: protesters, location, crowd size, …

  - Each event type has its own arguments
  - All "group" must be an Entity

6,217 Entities

25 Event Types

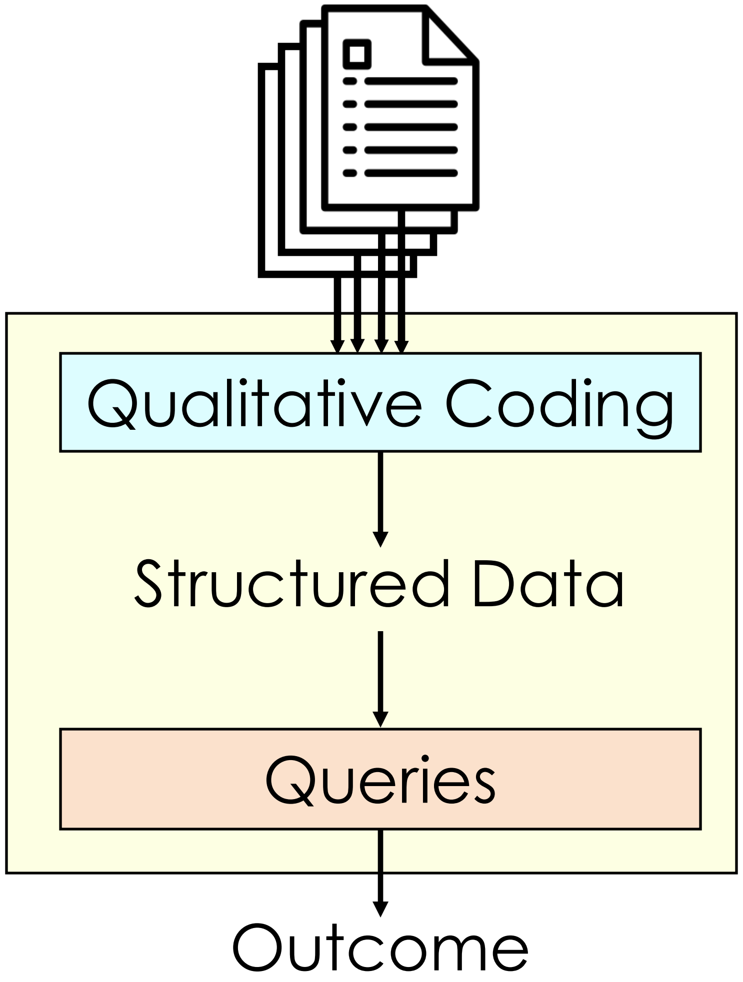
Problem 1. Event detection

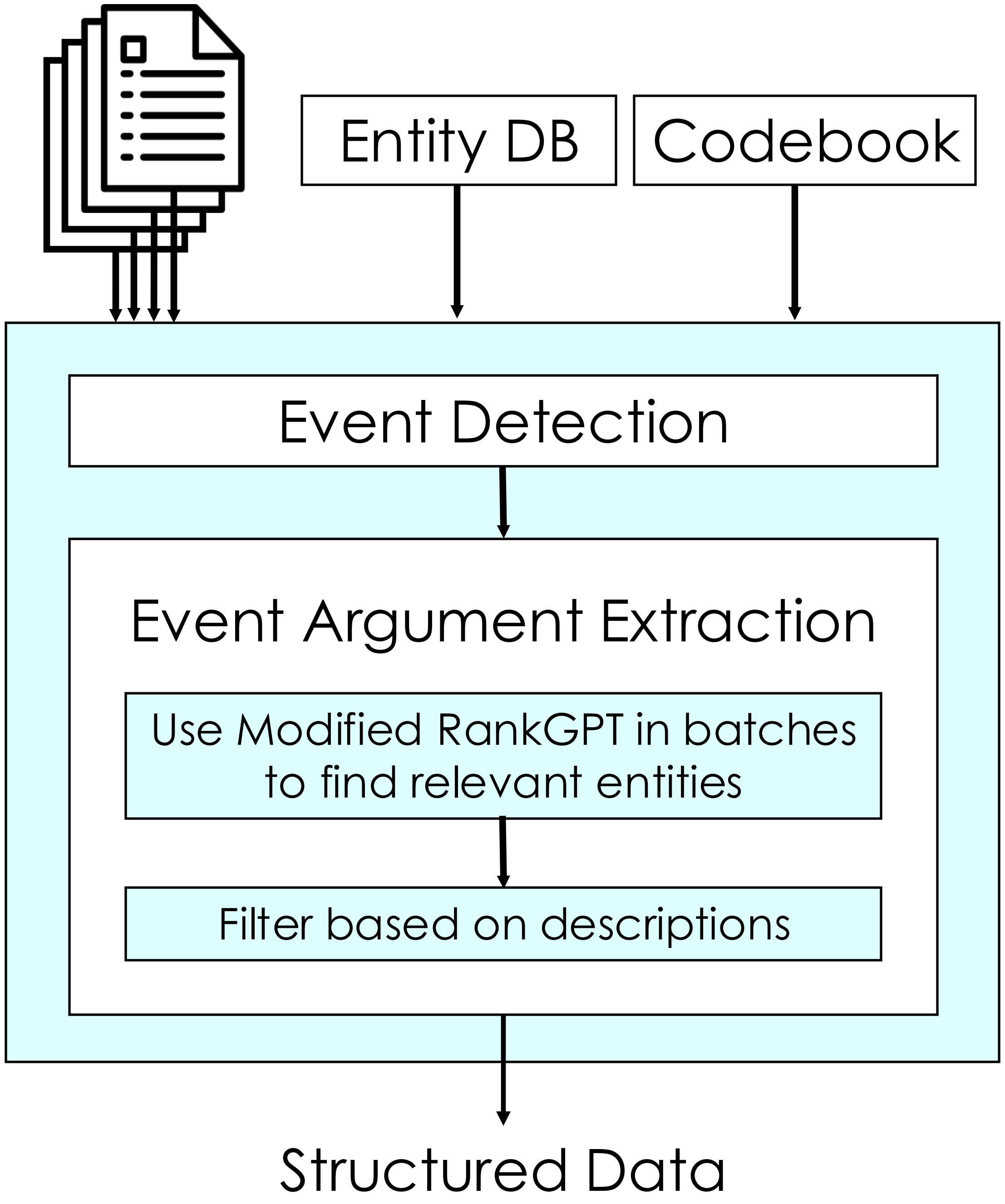
Event Arguments

Problem 2. Event argument extraction

Problem 3. Entity detection and linking

Zest:
0-Shot Qualitative Coder





Problem 1: Let’s Detect Event in this News Article

- From Indybay.org, a community news website
- 580 words

Tesla Fremont MLK Rally Protesting Racism, Union Busting

A rally on MLK Weekend was held at the massive Tesla Fremont assembly plant where 20,000 work. The action was called to protest the systemic racism and sexism by Elon Musk and his massive union busting drive. They also supported the Swedish striking Tesla mechanics who have been on strike for nearly 2 months.

…

<u>*https://www.indybay.org/newsitems/2024/01/15/18862100.php,*</u> Accessed 11/5/2024

Event Types in the Codebook

Event Type	Expert Description (summarized)
Peaceful Protest	Protestors do not engage in violence
Protest w/ Intervention	Protestors are faced with a physical attempt to disperse or suppress without serious or lethal injuries
Excessive Force Against Protesters	Protestors are targeted with lethal violence or violence resulting in serious injuries
Violent Demonstration	Protestors engage in violence and/or destructive activity
[20 more event types]	
Chemical Weapon	Chemical weapon is used in warfare, as listed as Schedule 1 of the Chemical Weapons Convention of 1993

Event-type Detection (ED) is a Classification Task

- Can be learnt by a lot of training data
- But annotation is costly
  - Lots of classes, lots of edge cases
  - Much more difficult than other classification tasks
  - Must be repeated for each codebook and domain

How **Can we use LLMs?**

Event Detection: Zero-shot LLMs

```
# instruction
Determining the best matching Event types for a given news article.

# input
First, here is the news article you need to analyze:

{{ article }}

Now, carefully review the annotation guidelines for various event types:

{% for e in event_types %}
  {{ e.name }}: {{ e.description }}
{% endfor %}

Your output should look like this:
<reasoning>
[Explain your reasoning for the event types you decide to include, and their order]
</reasoning>
<answer>
event_type_1 > event_type_2 > ...
</answer>
```

Result Preview: 80% (GPT-o4)

Problem 2. Event Argument Extraction (EAE)

- Now that we know the event type,
  zoom in on the relevant part of the codebook:

Event Argument	Expert Description (summarized)
Protestors	List of protestor groups or individuals involved in the protest
Location	Where the event happened
Crowd Size	Estimated size of the crowd. It can be an exact number, a range, or a qualitative description like 'small'.
Counter Protestors	Groups or entities engaged in counter protest, if any

Event Argument Extraction w/ LLMs – Attempt 1

```
# instruction
Extract event arguments from a given news article.

# input
First, here is the news article you need to analyze:
{{ article }}

Now, carefully review the event argument guidelines for the {{ event_type }}:

{% for ea in event_type.event_arguments %}
  {{ ea.name }}: {{ ea.description }}
{% endfor %}
```

LLMs Can Struggle to Follow Complex Instructions

- Types constraints, according to the codebook
  - E.g. crowd size should be a string, not number
- Nested event arguments
  - E.g. “Location” can have a “country” argument
- Relationship between different events

Idea: LLMs Know Python Data Structure Syntax!

- Use Python classes to represent the event signature
- Event types as classes
  - Event arguments as **typed** fields

Coding Guidelines as Class Definitions

```
class Protest(ACLEDEvent, ABC):
    """
    A "Protest" event is defined as an in-person public demonstration of
three or more participants in which the participants do not engage in
violence, though violence may be used against them … Excludes symbolic
public acts such as displays of flags or public prayers, legislative
protests, such as parliamentary walkouts or members of parliaments
staying silent, strikes …
    """
    location: Location = Field(..., description="Location where the
event takes place")
    crowd_size: Optional[str] = Field(..., description="Estimated size
of the crowd. It can be an exact number, a range, or a qualitative
description like 'small'.")
    protestors: List[str] = Field(..., description="List of protestor
groups or individuals involved in the protest")


class PeacefulProtest(Protest):
    """
    Used when demonstrators gather for a protest and do not engage in
violence or other forms of rioting activity, such as property
destruction, and are not met with any sort of violent intervention.
    """
    counter_protestors: List[str] = Field(..., description="Groups or
entities engaged in counter protest, if any")
```

```
class Location(BaseModel):
    """
    The most specific location for an event. Locations can be named
populated places, geostrategic locations, natural locations, or
neighborhoods of larger cities.
    In selected large cities with activity dispersed over many
neighborhoods, locations are further specified to predefined
subsections within a city. In such cases, City Name – District name
(e.g. Mosul – Old City) is recorded in "specific_location". If
information about the specific neighborhood/district is not known,
the location is recorded at the city level (e.g. Mosul).
    """

    country: str = Field(..., description="Normalized name of a
country, e.g. United States")
    address: str = Field(..., description="Full address or location
description including all geographic levels up to the neighborhood
level, including village/city, district, county, province, region,
country, if available. Exclude street names, buildings, and other
specific landmarks.")
```

Event Argument Extraction w/ LLMs Attempt 2

```
# instruction
Extract event arguments from a given news article.

# input
First, here is the news article you need to analyze:

{{ article }}

Now, carefully review the annotation guidelines for the {{ event_type }}:

<class definitions (on last slide)>
```

LLM is Good at Structured Output: Constrained Decoding for Structured Outputs

1. Convert Python class definitions to JSON schema
2. Convert the JSON Schema into a context-free grammar
3. Pass the prompt and JSON schema to the LLM
4. Decode the output from LLM, choose the most likely token
   that conforms to the grammar → this is the "constrained" part
5. Convert the decoded JSON object to the original Python object

- Some commercial LLM providers support it, e.g. OpenAI
- Many open source LLMs support it via
  - SGLang, Outlines, guidance

SGL outlines {guidance}

https://platform.openai.com/docs/guides/structured-outputs/introduction, Accessed 11/5/2024

Event Argument Extraction w/ LLM

```
PeacefulProtest(
    protestors=[
      "workers",
      "musicians",
      "ILWU",
      "CWA",
      "Teamsters Local 70",
      "PMWG",
      "former workers at Tesla",
      "Past ILWU president Brian McWilliams",
      "Nadya Williams"
      ],
    location=Location(
        country="United States",
        address="Fremont, California, United States"
    ),
    crowd_size="more than 100",
    counter_protestors=[],
)
```

Tesla Fremont MLK Rally Protesting Racism, Union Busting

Workers and musicians rallied on 1/13/24 at the Tesla assembly plant in Fremont, California on MLK weekend to protest Elon Musk's systemic racism and sexism at the Tesla assembly plant. They also protested the union busting and rallied in solidarity with striking Swedish Tesla service mechanics ... Workers from the ILWU, CWA, Teamsters Local 70, and PMWG spoke in solidarity, as well as former workers at Tesla. ...

Past ILWU president Brian McWilliams joined the Tesla MLK action and spoke ...

Rally participant Nadya Williams talked about her son, who is a Swedish American union organizer and is supporting the striking Swedish Tesla mechanics ...
By noon, there were more than hundred workers ...

Problem 3. Entity Detection & Linking

- "Entity" must refer to an item on the List
  - So you can link the different records to the same entity
- For example, in ACLED:
  - Generic entities like "Labor Group" annotated to aid analysis of labor issues
  - Specific Unions are monitored

```
protestors=[
    "workers",
    "musicians",
    "ILWU",
    "CWA",
    "Teamsters Local 70",
    "PMWG",
    "former workers at Tesla",
    "Past ILWU president Brian McWilliams",
    "Nadya Williams"
    ]
```

Abstractive Event Argument Extraction (AEAE) Problem

Given a database of entities, select all relevant to the event

Entity	Expert Description (summarized)
Labor Group	A collective entity composed of workers or trade unions that advocate for labor rights and interests …
AFL-CIO: American Federation of Labor and Congress of Industrial Organizations	The largest federation of unions in the United States. Encompasses various affiliated unions, such as the IW, IUPAT, **CWA** …
CWA: Communications Workers of America	A labor union representing workers in telecommunications, media, manufacturing, healthcare, and public service …
IBT: International Brotherhood of Teamsters	A labor union in the United States representing transportation and logistics workers. Operates through local chapters, like **Teamsters Local 70** …
ILWU: International Longshore and Warehouse Union	A labor union representing dock workers and other maritime and warehouse employees primarily in the United States …
[6212 more entities]	

AEAE Example

```
PeacefulProtest(
    protestors=[
        "Labor Group",
        "AFL-CIO: American Federation of Labor and Congress of
Industrial Organizations",
        "CWA: Communications Workers of America",
        "IBT: International Brotherhood of Teamsters",
        "ILWU: International Longshore and Warehouse Union",
    ],
    location=Location(
        country="United States",
        address="Fremont, California, United States"
    ),
    crowd_size="more than 100",
    counter_protestors=[],
)
```

Tesla Fremont MLK Rally Protesting Racism, Union Busting

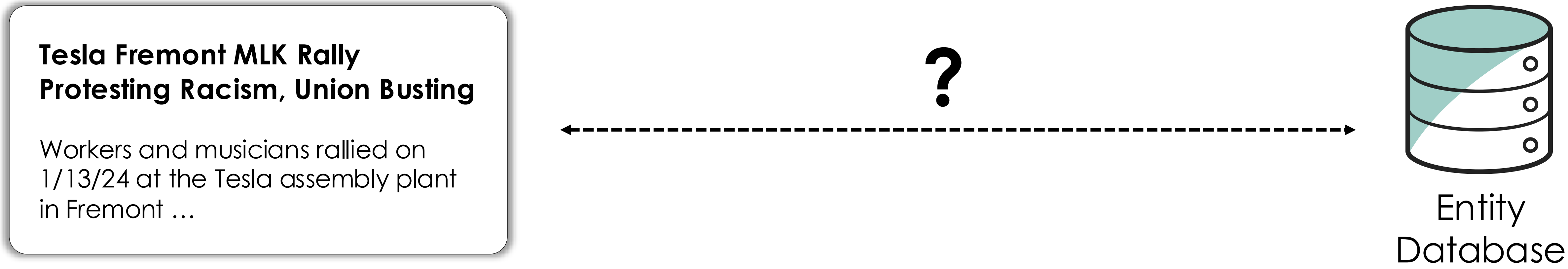
Workers and musicians rallied on 1/13/24 at the Tesla assembly plant in Fremont, California on MLK weekend to protest Elon Musk's systemic racism and sexism at the Tesla assembly plant. They also protested the union busting and rallied in solidarity with striking Swedish Tesla service mechanics ... Workers from the ILWU, CWA, Teamsters Local 70, and PMWG spoke in solidarity, as well as former workers at Tesla. ...

Past ILWU president Brian McWilliams joined the Tesla MLK action and spoke ...

Rally participant Nadya Williams talked about her son, who is a Swedish American union organizer and is supporting the striking Swedish Tesla mechanics ...
By noon, there were more than hundred workers ...

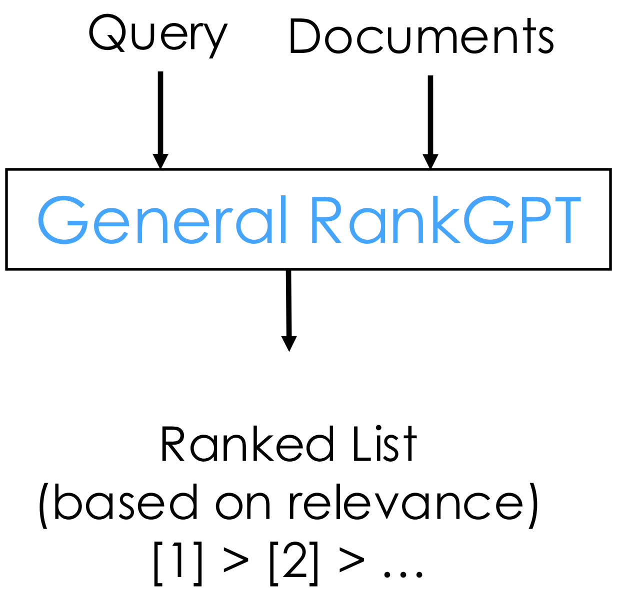
How To Perform AEAE?

- We need to connect the article to its entities
- Is like a retrieval task:
  - Given a document, retrieve the most relevant entities
  - But off-the-shelf retrievers don’t work well on this task



Use RankGPT to Retrieve Top Match

- Entities are curated by experts (< thousands entities)
- RankGPT [Sun et al.] proposes a 0-shot prompt for ranking items



RankGPT Prompt

```
# instruction
<query>
{{ query }}
</query>

{% for item in items %}
   [{{loop.index}}] {{ item }}
{% endfor %}

List the items from the most relevant to the
query to the list relevant.
Your output should be in the format [1] > [2]
> …
```

Is ChatGPT Good at Search? Investigating Large Language Models as Re-Ranking Agents, Sun et al, EMNLP 2023

Apply RankGPT to EAE

Document
Event Type
Event arguments’ description

Entity Database

Modified RankGPT

Only relevant entities

Role Assignment Prompt

Entities and their
Assigned Argument

Labor Group
…

Labor Group -> Protestors
…

```
# instruction
Your task is to select all entities involved in a news article from a provided list. An entity is an individual, group, or
organization involved in an event. This includes:
  - Organized armed groups with political purposes
  - Named entities
  - General terms describing participants like "Rioters", "Protestors", "Civilians", "Labor Group", etc.

# input
News article:
<article>
{{ article }}
</article>

The event you should focus on is the {{ event_type }} event, which happened in {{ country }}.

Guidelines:
1. Read the entire article carefully.
2. Identify groups, organizations, and individuals involved in the described events.
3. Note both specific names and generic terms used for participants.
4. Consider entities that may be implicitly involved.
5. For politicians, include the name of their political party or group as well, if available in the entity list.
6. Include both specific and generic entities when applicable (e.g., a political party leading a protest should be counted as two
entities: the party name and "Protestors"), if available in the entity list.
7. Include characteristics like ethnicity or religion as separate entities when mentioned (e.g., "Latin American Group" or
"Women"), if available in the entity list.
8. Err on the side of inclusion if unsure about an entity's involvement.

From the following list, select entities involved in the event described in this news article:
<entity_list>
{% for entity in potential_entities %}
[{{ loop.index }}] {{ entity }}
{% endfor %}
</entity_list>

Provide your answer within <entities> tags, listing one entity name per line:

<entities>
entity_name_1
```

Problem: The Database is Too large

- N = 6.2K entities won't fit in one prompt!
- Solution:
  1. Split database into groups of M_1 = 100 entities to find candidates
     - Will need N/M_1 LLM calls

Event Argument Extraction

Use Modified RankGPT in batches to find relevant entities

↓ candidates

Filter based on descriptions

Entity Database

→

Entity names 1-100

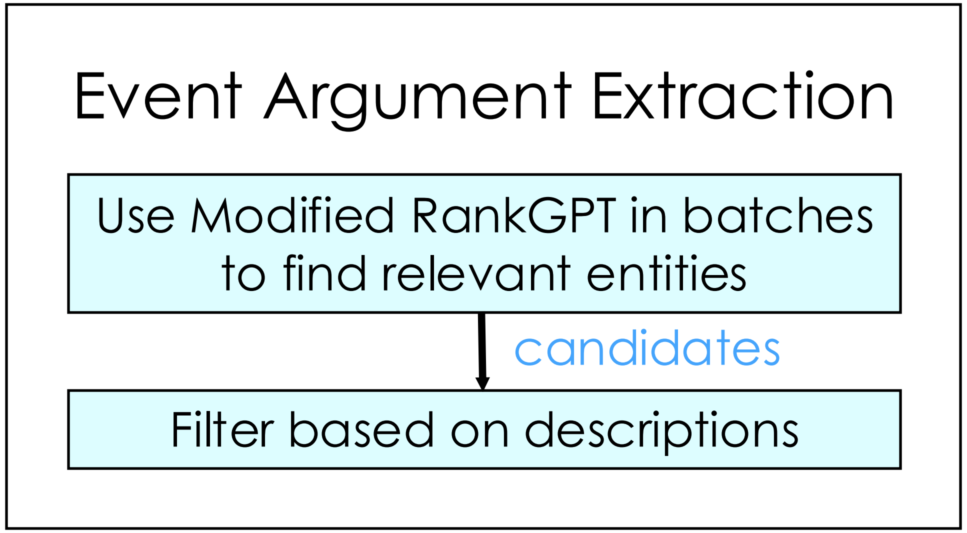
Entity names 101-200

…

→

Problem: The Database is Too large

- N = 6.2K entities won't fit in one prompt!
- Solution:
  1. Split database into groups of M_1 = 100 entities to find candidates
     - Will need N/M_1 LLM calls
  2. Split candidates into groups of M_2 = 10 entities
     - Prompt LLM to extract evidence from the article
     - Remove entities that don't have any evidence



Lecture Goals

- Motivation: Analysis across a document set
- ACLED: Manual qualitative coding
- Prior research on automated qualitative coding
- ZEST: an Automatic Qualitative Coder
- **LEMONADE: An AQC benchmark based on ACLED**
- Evaluation

LEMONADE

Large
Expert-annotated
Multilingual
Ontology-**N**ormalized
Abstractive
Dataset of
Events

- A cleaned version of ACLED event data
  - Training: Events from Jan to Jun of 2024
  - Validation/Test: Events from Jul 2024
  - 41,148 events in total

	Total	en	es	ar	fr	pt	ko	de	uk	my	it	tr	id	ru	fa	ne	zh
Train	17000	17000	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Dev	12074	1000	1000	1000	1000	1000	1000	1000	842	724	721	714	703	395	387	316	272
Test	12074	1000	1000	1000	1000	1000	1000	1000	842	724	721	714	703	395	387	316	272

LEMONADE statistics per language

The best-annotated dataset excerpted from a real-life dataset
with end-to-end abstract entity linking

Automatic Qualitative Coding (AQC)

- **ACLED is a major world-level effort**
  - Weekly effort by 200+ (part-time) world-wide researchers
- **Can AQC help ACLED expand coverage?**
  - Events, languages, countries
  - High-quality LEMONADE → effective fine-tuning
- **Can In-Context Learning based AQC lead to new analysis efforts?**
  - High-quality LEMONADE → compare in-context learning with effective fine-tuning

Lecture Goals

- Motivation: Analysis across a document set
- ACLED: Manual qualitative coding
- Prior research on automated qualitative coding
- ZEST: an Automatic Qualitative Coder
- LEMONADE: An AQC benchmark based on ACLED
- **Evaluation**

Evaluation: Event Detection

- Event Detection is a classification task → F1 metric
- Compare
  - LLaMA 3.1 8B (trained on 17K examples)
  - GPT-4o (0-shot)

Model	ED F1
LLaMA 3.1	87.3
GPT-4o	79.8

Other Tasks and Metrics

- **Event Detection (ED): F1**
- **Abstractive Event Argument Extraction (AEAE): F1**
  - Given the right Event Detection as input
  - Extract all Event Arguments, including Entities and their role
- **Abstractive Event Extraction (AEE): F1**
  - Do ED, then do AEAE using the outputs of ED
- **Abstractive Entity Detection and Linking (AEDL): F1**
  - Find the list of all relevant entities

Experiments on LEMONADE

- Fine-tuned 8B-parameter LLaMA 3.1
- Zero-shot system based on GPT-4o

	ED F_1	AEAE F_1	AEE F_1	AEDL F_1 (all)	AEDL F_1 (seen)	AEDL F_1 (unseen)
Supervised Models						
LLaMA 3.1 (8B)	87.3	77.3	67.5	68.6	80.9	14.1
+ *translation at test time*	88.5	**80.2**	71.0	69.9	82.0	17.2
Mistral NeMo (12B)	87.9	76.6	67.3	69.2	81.5	12.1
+ *translation at test time*	**89.6**	79.9	**71.6**	**71.3**	**83.0**	17.7
LLaMAX (8B)	88.3	76.7	67.7	68.3	80.5	13.3
+ *translation at test time*	88.1	79.0	69.6	70.3	82.2	16.3
LLaMA 2 (7B)	82.1	73.3	60.2	64.2	75.6	11.3
+ *translation at test time*	88.0	79.2	69.7	69.3	80.5	17.2
Zero-shot Models						
ZEST	79.8	71.7	57.2	54.0	55.3	**50.3**
- *entity linking*	79.8	54.8	43.7	8.5	18.4	0.2

Translation in test time: translate everything into English

Performance for Different Languages

AEE F_1 of two models on individual languages of the test set

Model	ko	zh	de	en	it	id	es	fa	pt	ne	fr	tr	ru	uk	ar	my
LLaMAX	81.3	79.3	78.5	76.7	76.6	76.4	72.3	67.7	66.1	65.9	65.6	64.0	63.7	62.9	48.9	41.2
Zest	57.4	67.0	71.0	51.7	70.4	60.1	60.3	63.8	52.4	51.9	61.5	60.0	61.4	54.7	40.9	43.8

ko: Korean, **zh**: Chinese, **de**: German, **en**: English, **id**: Indonesian, **es**: Spanish, **fa**: Farsi, **pt**: Portuguese, **ne**: Nepalese, **fr**: French, **tr**: Turkish, **ru**: Russian, uk: English (UK), **ar**: Arabic, **my**: Burmese (spoken in Myanmar)

Result Analysis (Supervised Models)

- **Event detection errors**
  - Some event qualities for multiple event type
  - ACLED codebook have rules to resolve ambiguities
    – need to be incorporated as expert "rules" ← Work in progress
- **Event argument extraction errors**
  - Abstractive entity detection and linking
    – need missing world knowledge ← Work in progress

- **Useful today for highlighting potential errors in hand annotation**

Result Analysis (Zero-shot Model)

- **Zest works better than expected**
  - Fine-tuning dataset size: 17000
  - 10 point drop in AEE F1: 57.2 vs. 67.7 F1 (ignoring test time translation)
- **Zest can handle unseen entities better**
  - 50 F1 vs. 14.1
  - (ignoring test time translation)
- **Struggles with edge cases in ED and AEDL**

Conclusion

- **Automatic Qualitative Coding (AQC)**
  - Has many applications
- **Lemonade: the best annotated event dataset excerpted from ACLED's real data**
  - With abstract event extraction
  - Fine-tuning reaches 67 F1
- **Zest: an AQC designed for real-life problems**
  - Contribution: abstract entity detection and linking
  - Promising zero-shot performance (10 points lower in F1)