MULTILINGUAL ABSTRACTIVE EVENT EXTRACTION FOR THE REAL WORLD

Anonymous authors
Paper under double-blind review

ABSTRACT

Event extraction (EE) is a valuable tool for making sense of large amounts of unstructured data, with a wide range of real-world applications, from studying disease outbreaks to monitoring political violence. Current EE systems rely on cumbersome mention-level annotations, and event arguments are frequently restricted to ungrounded spans of text, which hinders the aggregation and analysis of extracted events. In this paper, we define a new *abstractive* event extraction (AEE) task that moves away from the surface form and instead requires a deeper wholistic understanding of the input text. To support research in this direction, we release a new multilingual, expert-annotated event dataset called LEMONADE, which covers 16 languages, including several for which no event dataset currently exists. LEMONADE has $41,148$ events, and is based on the Armed Conflict Location and Event Data Project, which has been collecting and coding data on political violence around the globe for over a decade. We introduce a novel zero-shot AEE system ZEST that achieves a score of 57.2% F_1 on LEMONADE. With our supervised model that achieves 71.6% F_1, they represent strong baselines for this new dataset.

1 INTRODUCTION

Event extraction (EE) is an important tool for studying the real world. Its applications span a wide range of fields, from social sciences (Zubiaga et al., 2014) to biomedicine (Lybarger et al., 2021; Kim et al., 2003). It is used for early detection and tracking of disease outbreaks (Parekh et al., 2024; Consoli et al., 2024; Min et al., 2021), monitoring cybersecurity threats (Satyapanich et al., 2020), studying political conflicts (Hu et al., 2022), protests (Radford, 2020; Zhukov et al., 2019; Hürriyetoğlu et al., 2022a; Zavarella et al., 2022), and crime (Mostafazadeh Davani et al., 2019). Because of its costly annotation process, automated EE systems are highly desirable.

In AI research, automated event extraction has been an extensively studied topic in information extraction (Ji & Grishman, 2008). However, the resulting EE systems have several shortcomings that keep them from real-world applications (Hürriyetoğlu, 2021; Hürriyetoğlu et al., 2022b; 2023; 2024a). Monitoring socio-political developments perhaps best exemplifies the requirements of event extraction (automated or not) in the real world.

Entity Normalization and Linking One of the main uses of event data is trend discovery and aggregate reporting (Li et al., 2019a; 2020b; 2021a; Reddy et al., 2023). Traditional EE systems, which construct extractions based on text spans (Huang et al., 2024), are ill-suited for this purpose. This is especially important for entity arguments; most EE systems either do not link entities, or use tools that link to Wikidata (Wen et al., 2021) or Wikipedia (Li et al., 2019a; 2020a), which do not necessarily match the expectations of the domain, leading to the need for domain-specific entity datasets and systems (Wei et al., 2016). As such, an EE system should facilitate event argument normalization, and support linking entities to a any provided entity database.

High Demand for Annotation Quality Even manual annotation of events is challenging, and poor annotation quality is especially detrimental as it contributes to biased inferences in high-impact policy decisions such as international peacemaking efforts (Andrea Ruggeri & Dorussen, 2011). This often necessitates expert annotations instead of crowdsourcing (Raleigh et al., 2010; Caselli & Huang, 2012). As such, automatic EE systems should be built and evaluated using high quality data.

Multilinguality To study the real world, we often need a *global* view, which necessitates support for a wide range of languages, especially low-resource ones, as for example much of the political analysis of outbreaks and conflicts is focused on the global south and the international setting. Existing event datasets only cover a few languages such as English and Chinese, therefore, EE systems are not properly evaluated on how well they can be used to study global phenomena.

Flexible Schema and Ontology It is important to support custom schemas and entity lists. Many codebooks have been developed for events over many decades of work (Azar, 1980; McClelland, 1978; Walker et al., 2006; Gerner et al., 2008; Walker et al., 2006; Halterman et al., 2023a; Tracey et al., 2022; Duruşan et al., 2022). Oftentimes, scholars define a new domain-specific schema for the phenomena they want to study de Mesquita et al. (2015). While recent work in zero-shot information extraction has made advancements in this direction Sainz et al. (2024), they do not generalize well to arbitrarily varied schemas (Section 5.2).

Intermediate Annotations are not Available In EE literature, the task, datasets and systems are all typically divide into several parts (Huang et al., 2024), each requiring cumbersome span-level annotations: 1) event trigger identification, 2) event trigger classification, 3) event argument identification, 4) event argument classification, 5) entity detection, 6) entity coreference resolution 7) entity linking, and 8) event coreference resolution. Different works either work on a subset of these tasks, or lump them together under the names like event detection (1 and 2), or event argument extraction (3 and 4). Even document-level EE (Tong et al., 2022) relies on span-based intermediate annotations for the task. These intermediate annotations add to the cost of obtaining data for a new domain, and make high quality annotations even more challenging.

In summary, automatic EE in the real world remains challenging. To study a new phenomena (or and old phenomena from a new angle), we need high quality data, often multilingual and with normalized entities across different languages. As an example of the level of effort required, Armed Conflict Location and Event Data (ACLED) (Raleigh et al., 2010; 2023) is annotated by a team of 200 researchers from around the globe (Sam Jones, 2022). To make matters more challenging, off-the-shelf tools like entity linkers that work against Wikidata are not applicable to many domains (Wei et al., 2016). These limitations have remained largely unchanged even with the recent use of large language models (LLMs) and in-context learning in EE (Wang et al., 2023; Sainz et al., 2024).

In this paper, we attempt to bridge this gap between the real-world requirements and EE research by making a real event dataset available, and by evaluating the use of NLP technology to assist in real world EE.

The Abstractive EE (AEE) Problem. Aiming to create a useful tool for real-world EE, we formulate the AEE problem. The distinguishing factor in AEE is that it moves away from the surface form of the text, and focuses on grounding events on a predefined ontology like an entity database, or categorical event arguments [1]. We define the AEE problem as follows:

Definition 1

We define event extraction codebook $C = (T, \mathcal{D}, S)$ where

- T is the set of possible event types,
- Each $D \in \mathcal{D}$ is a domain such as integers, real numbers, or a set of known entities,
- A list of event signatures $S = [(t_1, a_{1,1}, \ldots, a_{1,n_1}), ...]$, where n_i is the number of arguments for event type t_i, and $a_{i,j}$ is an argument with domain $D_{i,j} \in \mathcal{D}$

Definition 2

The *Abstractive Event Extraction* (AEE) problem is: given codebook $C = (T, \mathcal{D}, S)$ and writing $w \in W$, extract abstractive event AEE$(w, C) = (t_i, v_1, ..., v_{n_i})$ which is the main event Tong et al. (2022) in w, where t_i is the ith event type in T, $v_j \in D_{i,j}$ and n_i is the number of arguments for event type t_i .

[1]The term *abstractive* has been used in other NLP tasks like OpenIE (Pei et al., 2023) and summarization (Radev et al., 2002) to refer to the concept of moving away from the surface form.

In the example in Figure 1, t_i = MobViolence $\in T$, the first two arguments, group_1 and group_2 represent the two sides in the violence, with $D_{i,1}, D_{i,2}$ being the set of all subsets of possible entities from the event database, the third argument is a location, and domains of the last two arguments, $D_{i,4}, D_{i,5}$, are both booleans.

In AEE, we remove the limitation for arguments to be spans, or explicitly mentioned in the text at all. In addition to the abovementioned benefits, this also enables the annotation of *implicit* event arguments. For instance in Figure 1, the higher-level entities like "Dalit Caste Group" require domain-specific knowledge (the caste system in India in this example), which is provided as a descriptions in the entity database.

The LEMONADE Dataset. We present an event dataset for the AEE task called LEMONADE (**L**arge **E**xpert-annotated **M**ultilingual **O**ntology-**N**ormalized **A**bstractive **D**ataset of **E**vents). The dataset is extracted from the high-quality data annotated by experts at ACLED. This data has been used by international organizations like The United Nation's International Organization for Migration, The International Rescue Committee and The European Commission for tracking and predicting forced displacements and evaluating humanitarian efforts (ACLED, 2023).

Solving the AEE Problem In this paper, we study the following questions:

1. Given a high-quality AEE training dataset, can we perform AEE effectively?
2. It is costly to create a large high-quality AEE training dataset for new domains. Is it possible to create a zero-shot model for AEE?

The contributions of this paper include:

- A new expert-annotated dataset called LEMONADE. It includes $41,148$ events covering 16 languages, including several languages like Indonesian, Burmese and Nepali that were not previously studied for events in an academic setting. LEMONADE has many entities that do not have Wikidata or Wikipedia entries, making it especially challenging and a suitable testbed for zero-shot entity linking systems.
- Our supervised AEE model achieves 71.6% F_1 on LEMONADE, establishing a strong baseline.
- We propose ZEST, a novel zero-shot system for AEE. To handle the full complexity of the real-world AEE problem, we decompose the problem into manageable subproblems; of note is the novel zero-shot entity linking component. The zero-shot ZEST achieves 57.2% on the LEMONADE, which is 13.5% better than existing zero-shot baselines.

2 RELATED WORK

The task of Event Extraction aims to extract events and their arguments from a given context. The Message Understanding Conferences (MUC) in the 1990s (Grishman & Sundheim, 1996) were one of the first endeavors at building automated EE systems (Anderson et al., 2012). The datasets prepared for MUC pioneered text spans as the unit of some system outputs. Today's EE research is based on the task formulation of the ACE05 project (Walker et al., 2006), which divides the task into subtask at the sentence level with span-based intermediate annotations (Walker et al., 2006). Li et al. (2021b) extended EE to allow for arguments of an event to be from surrounding sentences, and Li et al. (2021b) introduce the concept of "most informative span" for arguments. Tong et al. (2022) introduced the DocEE dataset, where event arguments are scattered across the document, fully realizing EE as a *document-level* task.

EE has been extensively studied in the AI community (Ji & Grishman, 2008; Liao & Grishman, 2011; Chen et al., 2015; Liu et al., 2018; Yang et al., 2019; Zhu et al., 2024b; Ren et al., 2024; Lai, 2022; Li et al., 2022; Zhou et al., 2020). Previous work has employed a variety of approaches including graph-based modeling, which leverages structured relationships within data (Dutta et al., 2021; Lai et al., 2020; Zhang et al., 2020) and language modeling (He et al., 2015; Michael et al., 2018; Li et al., 2019b; Du & Cardie, 2020). Furthermore, joint modeling techniques (Nguyen et al., 2022; Hsu et al., 2022; Zhang & Ji, 2021) sometimes dubbed end-to-end models (Zheng et al.,

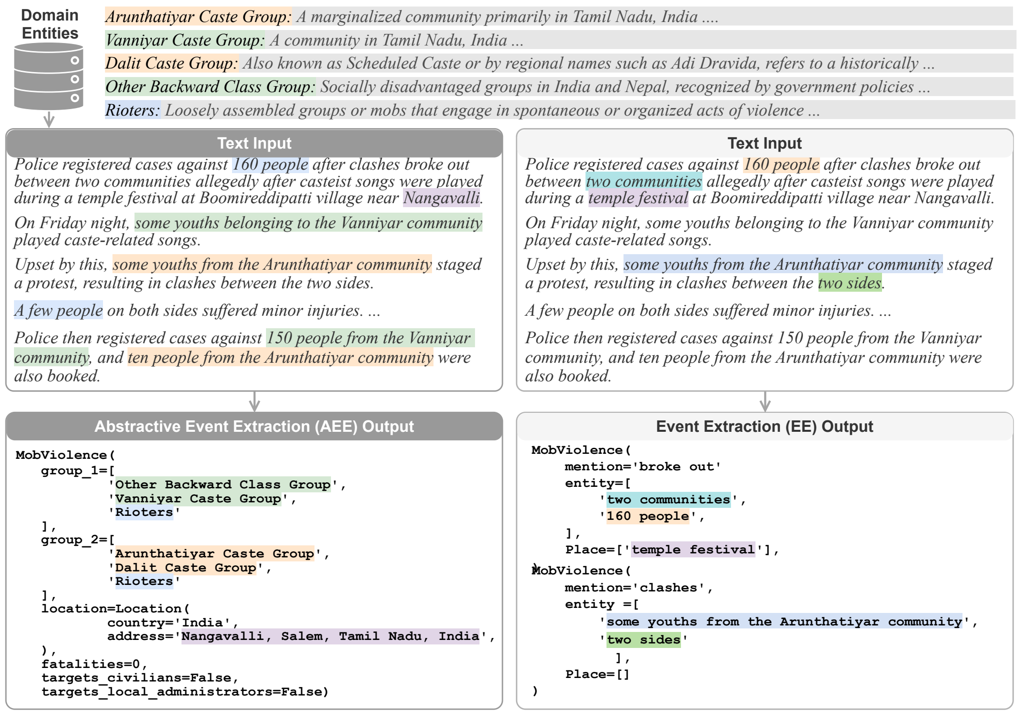

Figure 1: An example of a "Mob Violence" event from the LEMONADE dataset showing the many significant differences between AEE and EE on the same text input. For AEE, entities must be matched to one of the entities in the given domain, whereas EE annotations refer to the entities as span in the text. AEE identifies that it is a single event, whereas EE classifies it as two, with no way to annotate the two sides of the conflict consistently across them. AEE notes the two clashing groups explicitly. Group 1 includes not just "Vanniyar Caste Group" but "Other Backward Class Group", the larger group that the Vanniyar caste belongs to under the Indian government definition, as well as "Rioters" to indicate the presence of rioters. Similarly, group 2 includes not just "Arunthatiyar Caste Group", but the larger group "Dalit Caste Group" and the generic rioters group. The larger group information requires knowledge beyond what is in the text input; this is important to annotate given the known rivalries between the groups. Furthermore, the location information in AEE is much more precise than that of EE, enabling spatial analysis.

2019), integrate multiple EE subtasks to improve extraction accuracy. With the recent advances in generative language models, more research has focused on applying generative methods on event extraction (Shi et al., 2023; Anantheswaran et al., 2023; Li et al., 2021b; Lu et al., 2021), culminating in the use of LLMs (Xu et al., 2023; Wang et al., 2023; Qi et al., 2024). We note that AEE is different from end-to-end approaches, in that it does not rely on intermediate annotations of auxiliary tasks.

Most EE datasets focus on English, and to a lesser extent Chinese (Zhu et al., 2024a; Ren et al., 2024; Walker et al., 2006). Event extraction datasets for other languages include BKEE (Nguyen et al., 2024) for Vietnamese, InDEE-2019 (Maheshwari et al., 2019) for 5 Indic languages, MEE (Pouran Ben Veyseh et al., 2022) for Portuguese, Spanish, Polish, Turkish, Hindi, Japanese and Korean, Zavarella et al. (2014) for Bulgarian, Romanian and Turkish, and Balali et al. (2022) for Farsi.

While there are several socio-political event databases that use automated tools for extraction (Leetaru & Schrodt, 2013; Hallberg, 2012), manual annotation remains the gold standard.

Event extraction in the socio-political domain has long been an important research theme (Raleigh et al., 2010; Chenoweth & Lewis, 2013; Weidmann & Rød, 2019; Kriesi et al., 2019; Hürriyetoğlu et al., 2024b). A line of recent work uses language models to detect socio-political events with nuanced contextual understanding (Tanev, 2024; Tanev & De Longueville, 2023; Mehta et al., 2022; Slavcheva et al., 2023). Since data scarcity is a critical issue in socio-political event extraction, finding innovative data utilization strategies has also become a focus (Loerakker et al., 2024; Bakker et al., 2024; Mutlu & Hürriyetoğlu, 2023; DeLucia et al., 2023; Raj et al., 2022).

3 LEMONADE, A MULTILINGUAL AEE DATASET FOR THE REAL WORLD

LEMONADE is an expert-annotated event dataset covering 16 languages: English, Spanish, Arabic, French, Portuguese, Korean, German, Ukrainian, Burmese, Italian, Turkish, Indonesian, Russian, Farsi, Nepali, and Chinese. These languages are selected for their typological diversity (Clark et al., 2020), and span several high and low resource languages. To the best of our knowledge, this is the first event extraction dataset for Burmese, Indonesian and Nepali, and covers the most number of languages than any other event dataset by far.

In event extraction, intermediate annotations like event mentions and entity mentions are expected from datasets and systems (Liu et al., 2021). One event can be mentioned multiple times in the document, and those are called *event coreferences*. LEMONADE, on the other hand, following AEE, does away with annotating entity spans and coreferences, and event mentions and coreferences. Instead, it focuses on actually reporting the event that the document describes.

LEMONADE is based on the Armed Conflict Location and Event Data (ACLED) (Raleigh et al., 2010). Originally published in 2010, ACLED focused on civil war, subnational and transnational violent events in 50 unstable countries, it has since expanded to track more types of political violence event, as well as civil unrest events, in 243 countries and territories in 100 languages in near real-time (Sam Jones, 2022; ACLED, 2023). We chose this as our data source because in addition to the wide language coverage, it has high-quality expert annotations, mitigating quality issues present in many NLP datasets (Campagna et al., 2022).

In the rest of this section, we describe the process of creating LEMONADE.

We preprocess the ACLED data with the goal of transforming it into a format that is more amenable for AI models, while keeping as much of the information as possible. The main challenge is to ensure annotations only contain information that can be extracted or inferred from the input. The steps taken involve data cleaning and reannotation of certain event arguments. The general process was automated as much as possible, and involved spot-checks and several rounds of improvements from two authors of this paper.

ACLED Annotation and Review Process. We start from the publicly available expert annotations of ACLED. ACLED annotations are done by a group of around 200 experts and is updated on a weekly basis. It sources *writings* from news media, international organizations, NGO and security reports, and local partner organizations and select social media channels. It annotates one event per writing, the *main* event excluding historical events that are typically mentioned in writings to provide more context. These writings go through a multi-step review and quality assurance process (ACLED, 2020). The annotation of events is done at a regional level (e.g. the Middle East, Africa etc.) by experts of those regions. These experts have local language skills and knowledge about regional conflicts, and many live within the country they cover. The annotations are then merged by a research manager who reviews these data for inter-coder reliability across the region. Researchers use an annotation tool that provides them with the up-to-date list of entities and locations, and communicate with each other to clarify difficult annotation decisions. After merging regional data, another round of manual reviewing is performed by another expert.

There are 25 politically significant event types covering battles, protests, riots, violence against civilians, political agreements, arrests and more. Appendix D shows the full list of event types and the arguments of each one.

Data Filtering and Cleaning We obtain all events from the first 7 months of 2024. Overall, this includes $112,885$ events, each paired with a writing and an annotated event. After analyzing the data, we realized that many social media posts in the data are accompanied by an image (e.g. protest fliers), and the text alone is not enough to annotate the event. Therefore, we exclude social media posts. We also remove the 1% longest and shortest writings, because very short ones (often from local partner organizations' reports) do not include enough context for annotating the event, and very long texts are often a combination of multiple news articles. This leaves us with 90,035 events sourced mainly from mostly news articles. A number of ACLED events include multiple writings and annotations, each one covering one aspect of the event, for example, a national protest that occurs in multiple cities. We keep one of each event, and are left with $63,217$ events. We further limit the data to languages that have at least 500 events. We obtain the writings from the provided URLs, and clean them by removing advertisements etc. using an LLM prompt.

Entity Database ACLED annotates entities involved in each event. We provide a database of 6217 entities that appear in ACLED events in 2024. In each event, entity arguments have a small subset of this database as their value. This list contains both generic entities (Halterman et al., 2023b) like "Rioters", "Women", "Students", and specific entities like "Vanniyar Caste Group".

Often, domain knowledge is required for entity linking in specialized domains. The example in Figure 1 demonstrates this. There are entities that are explicit mentioned in the source article and need to be linked to the database, and there are entities whose role in the event is *implicit*, or are annotated because of their relationship with an explicit entity.

While it is possible to learn entities of a domain with enough data, we want LEMONADE to enable research on zero-shot entity linking in this challenging setting. Therefore, to make domain knowledge available to models in a realistic way, we also provide a one-paragraph description for each entity. These descriptions are meant to provide entity linking models with enough context and domain knowledge to understand domain entities, especially the long tail (Mallen et al., 2023).

Location. Location is a crucial event argument for conflict events. In ACLED, the country and up to three subnational administrative levels are annotated (ACLED, 2023). In cases where an event happens in an unknown location within a larger geographic area, or near a city or border, the closest location is used as the location. In rare cases, other sources like maps are used to pinpoint the exact location of an event. There are two issues with this approach when used for building or evaluating AI models. First, because the annotations contain locations that are not inferrable from the writing, this would encourage models to *hallucinate* a location. Second, it puts the burden of knowing the location hierarchy (e.g. which town is in which province) on the shoulders of the AI model. For these reasons, we provide a simpler definition for location, and reannotate the location argument to match this definition: "*The location argument is the most specific place that is supported by the writing*".

For reannotation, we use the original ACLED location annotations to consult the OpenStreetMap geographic database (OpenStreetMap contributors, 2017) to find the full hierarchy of location above the neighborhood level for each event. We then start from the lowest location level and remove the items that are not supported by the writing, until we reach one that is. We then keep that location and all levels above that. A carefully designed LLM prompt was used for this last stage. The final location arguments were spot-checked by the authors, and 97% of them were correct according to the above definition. The Location argument in Figure 1 shows an example output of this reannotation process. In addition, during evaluation (Section 5.1), we first use the same geographic database to normalize locations, in case the AI model predictions have slight differences such as different spelling of town names, or a missing province name when the town name is extracted correctly.

Schematization ACLED uses the same event argument roles for all event types, resulting in some argument roles being always empty for some event types or the argument names being too generic. we define separate event argument roles for each event type. For example, we remove "fatalities" argument from "Peaceful Protest" and rename "actor 1" to "Abductor" for the "Abduction or forced disappearance" event type. We also provide a short description for each event type, and expert descriptions for each event argument, to facilitate the development of zero-shot models.

Following the recent trend in event extraction, we use Python code to represent annotations. This has been shown to improve the performance of various supervised (Sainz et al., 2024) and few-shot (Wang et al., 2023) models because it makes the labels closer to the code data many language models have been pre-trained on. Furthermore, this enables the use of constrained decoding (Rabinovich et al., 2017; Willard & Louf, 2023) algorithms to eliminate malformed outputs. Appendix D presents the full schema for LEMONADE.

Data splits We provide validation and test sets in 16 languages, and a large training set in English. The data split is across time, meaning that the events in the training set are from the first 6 months of 2024, and the events in validation and tests sets are from July 2024. This mimics the real-world setting where the distribution of events and entities might change over time. Because of this

Table 1: LEMONADE statistics per language.

	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

split, 22.1% of entities in the validation and test sets are not seen in the training set. The split between validation and test sets is random. Table 1 shows the language statistics of LEMONADE, and Appendix A contains event type and geographical distribution of the dataset.

4 ZEST: A ZERO-SHOT AEE MODEL

LEMONADE is the rare case where a large high quality training set is available, but that is not the case for many scenarios. In this paper we want to leverage LEMONADE to understand how we can tackle AEE, without requiring expert annotations for training. We assume no access to training data in any language, and that only the information about the schema and the domain is provided in the form the event ontology, and the entity database.

For this, we present a zero-shot system called ZEST. ZEST uses zero-shot in-context learning (i.e. only instructions). The inputs to AEE, writing w and codebook C, can be really long, with each event type having its argument signature. It is ineffective, if we present the LLM with the entire codebook. Our preliminary experiments showed that adding few-shot examples is inadequate, perhaps also due to the large size of w.

To address the complexity of AEE, we break it down into 3 simpler tasks that are more amenable to in-context learning:

1. *Event Detection* (ED) finds the abstractive event type;
2. *Abstractive Entity Detection and Linking* (EDL) finds a subset of the entity database involved in the abstractive event and assign them to the correct event argument;
3. *Abstractive Event Argument Extraction* (EAE) finds the event arguments for non-entity arguments, given the event type.

Note that EDL and EAE are handled differently from each other in ZEST, because the very large size of the entity domain adds more challenges that a zero-shot system needs to handle. Formally:

Definition 3

Given codebook $C = (T, \mathcal{D}, S)$ and writing $w \in W$,

$$\text{ED} = t, \text{where AEE}(w, C) = (t, \ldots)$$

$$\text{EDL}(w, C, t) = V, \text{where AEE}(w, C) = (t, v_1, \ldots) \text{ and } v_i \in V$$

$$\text{EAE}(w, C, t) = [v_1, \ldots], \text{where AEE}(w, C) = (t, v_1, \ldots)$$

ZEST Event detection (ED) Given that the list of event types (T) is relatively small (25 in the case of LEMONADE), event detection can be done as a zero-shot in-context learning task. The prompt (Table 5) includes the input writing w and a list of event types and their descriptions. The task is to return the most likely event type t. We use chain-of-thought (Wei et al., 2023) for this prompt.

ZEST Entity Detection and Linking (EDL) Once the event type is determined, the next step is to narrow down the list of possible entities that are involved in the event.

We found that in-context learning cannot handle the large number of entities (6217 in the case of LEMONADE) in the AEE task if they are presented in one prompt. Therefore, we tackle this in two stages: the first narrows down the number of candidate entities and the second stage further more closely filters down the set.

We divide the list of all possible entities into groups of N. We use a simple zero-shot prompt (Table 6) that given w, t and all entities in each group, removes the irrelevant entities. In practice, we find that a wide range of values for N works well, and we choose $N = 63$ (i.e. 100 groups in the case of LEMONADE) in our experiments.

Given the w and the smaller list of entities and their description, the next step uses another prompt (Table 7) to find evidence of each entity's involvement in the event and remove the ones for which we cannot find any evidence. We find that this formulation is especially helpful in identifying implicit entities.

The last step is to match each entity with its correct event argument (e.g. is an entity the "victim" or the "perpetrator" of an "Attack" event?). For this, we use another prompt (Table 8) that given a list of entities and event arguments, outputs a mapping between the two.

ZEST Event Argument Extraction Given the identified event type and entity arguments, we now extract all the other arguments using an approach similar to Wang et al. (2023). This is done using a prompt (Table 9) that given w and the event type signature for t, outputs all non-entity argument values.

5 EXPERIMENTS AND RESULTS

In addition to the zero-shot setting, we also measure the performance of the system separately in English and non-English languages (i.e. zero-shot cross-lingual generalization), and unseen actors (i.e. zero-shot generalization to unseen actors).

5.1 METRICS

To evaluate a predicted event against a gold event from LEMONADE, we first normalize the location arguments using a lookup in the OpenStreetMap geographic database. We then use simple string equality to calculate precision, recall and micro-averaged F_1 (Manning et al., 2008).

For ED, we compare the predicted event type against the gold event type, and report the micro averaged **ED** F_1. For EAE, we force the gold event type as the first part of the model output, and have it generate event arguments and their values $\{(a'_1, v'_1), ...\}$. We then consider this set as the returned result, and calculate its precision, recall and F_1 against the gold $\{(a_1, v_1), ...\}$ and report **EAE** F_1. In other words, two arguments are considered equal if their argument *and* values match.

We define and choose **AEE** F_1 as our main metric, which is similar to EAE F_1, except that if the predicted event type is incorrect, all arguments are considered incorrect, contributing to both false positives and false negatives in the calculation of F_1.

For entities, we report **EDL** F_1, which is the result of comparing the entity IDs between prediction and gold. Note that EDL F_1 ignores the argument type a. Additionally, we report EDL on two interesting subsets of entities: entities that have been *seen* in LEMONADE's training set, and those that are *unseen*.

5.2 SETUP

Supervised Setting If enough training data is available, we show that simply modeling the task as a sequence-to-sequence task is effective: the model is given w as input, and is trained to predict the full Python code representing the event. For this setting, we fine-tune several language models on the English LEMONADE training set. We use the 8-billion parameter version of LLaMA 3.1 (Dubey et al., 2024) for its strong performance in multilingual benchmarks. We also include LLaMAX (Lu et al., 2024), which extends LLaMA 3 to more than 100 languages by continual pre-training and the 12B parameter model `Mistral-Nemo-Base-2407` for its tokenizer's better support of non-Latin scripts. For comparison, we also include the 7-billion parameter version of LLaMA 2 (Touvron et al., 2023), which has not been specifically trained for non-English languages, though its pre-training data contains a small amount. The base (non-instruction-tuned) versions of all models are used.

We also experiment with *translation at test time* (Moradshahi et al., 2020), by translating all w in the test/dev sets into English using GPT-4o. This way, the supervised AEE model receives English text as input at inference time, which matches its training data more closely.

Zero-shot Setting For all zero-shot experiments, we use GPT-4o version `gpt-4o-2024-08-06`. We measure the impact of the zero-shot EDL of ZEST separately. We use constrained decoding when generating Python code for all settings, so the outputs are always syntactically valid, e.g. the event arguments are valid for the predicted event type. The most promising zero-shot baseline from the EE literature is GoLLIE (Sainz et al., 2024), given that it claims to support flexible schemas. However, while we were able to reproduce its results on the datasets they experimented with, the outputs were poor when evaluated with even a small change to the "Location" field. We believe this is due to the limited diversity in event schemas in its training data.

5.3 OVERALL RESULTS

Table 2 shows the result of our supervised and zero-shot systems on the LEMONADE test set, averaged over the 16 languages. LLaMA 3.1, LLaMAX and Mistral NeMo perform similarly, all within 0.2% of each other in the AEE F_1 metric. The added language support in LLaMAX has minimal effect. We attribute this to the fact that in LEMONADE, all outputs are normalized (and therefore in English), so the models have an easier task generalizing to new languages. Translating the documents to English, improves the AEE F_1 between 1.9% and 4.3%. The LLaMA 2 model which has not gone through special multilingual pre-training, on the other hand, sees the most benefit from *translation at test time*, with an improvement of 9.5% in AEE F_1.

As for our zero-shot system ZEST, it is 14.4% and 10.5% behind the best supervised (Mistral Nemo + translation) and the best supervised model without translation (LLaMAX) in terms of AEE F_1. The majority of this gap comes from ED (9.8% and 8.3% gap), while EAE is closer (8.2% and 5.0% lower). One area that ZEST shines, is in entity linking accuracy. Specifically, it adds 45.5% over the baseline of directly generating entities with LLM, and outperforms the supervised models in the unseen entity setting by at least 32.6% When training data is available for entities, however, supervised models significantly outperform ZEST.

Table 2: Results of our zero-shot and supervised systems on the test set of LEMONADE. Numbers are averages over all 16 languages. The highest number for each metric is in **bold**.

	ED F_1	EAE F_1	AEE F_1	EDL F_1 (all)	EDL F_1 (seen)	EDL 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

5.4 PER-LANGUAGE RESULTS

We take a closer look at the performance of the best cross-lingual model (LLaMAX without translation), and ZEST in each individual language. Table 3 shows per-language results on the LEMONADE test set. The largest gap between the supervised model and ZEST is in English (25% in AEE F_1), which is reasonable given the training data for LLaMAX is in English. We provide the language acronym mapping in Appendix.

Our analysis of the outputs show that the variance between languages is mainly due to the different distributions of event types. For instance, in politically stable countries (where writings in Korean, Italian, Chinese and German languages come from), almost all event types are of "Protest" type, and there are no battles or remote violence reported, and we observe that ED score for supervised methods is really high. Overall, given the abstractive nature of the task, and the fact that the gold annotations are normalized and in English, the effect of cross-lingual capabilities of the model becomes less influential relative to extractive EE.

ZEST outperform the supervised model in Burmese (my). This language, widely spoken in Myanmar, has a wide range of event types in LEMONADE, and due to its low-resource nature, is quite challenging in the cross-lingual setting. Russian, Farsi, Turkish and French are other languages where the gap is relatively small.

Table 3: AEE F_1 of two models on individual languages of the LEMONADE test set.

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

6 CONCLUSIONS

This paper introduces the task of abstractive event extraction (AEE), which more closely matches the requirements of event extraction for real-world applications. We have derived a large high-quality dataset for the AEE task, in 16 different languages, from the expert-annotated data created by ACLED.

We introduced ZEST, a novel zero-shot AEE system, that achieves 57.2%. With our supervised model that achieves 71.6% F_1, they represent strong baselines for this new dataset.

Reaching 71.6% with supervised learning, our system can be helpful to human annotators by providing them with the first draft to accelerate the annotation task. Furthermore, errors do occur in human-annotated data. The automatically generated results can be used to double check human annotations. During the error analysis of ZEST for example, we discovered a few missing entities in the manual annotations. In contrast, we note that the original EE formulation that refers to entities as spans in the text is not useful for event analysts, nor can it be used to help human annotators.

ETHICS STATEMENT

No human subjects were involved in this study. We will release LEMONADE in accordance with the ACLED Terms of Use. ACLED data do not contain personally identifiable information (e.g. names of individuals or mobile device IDs), and cannot be used to track individuals. No crowdsourcing was performed as part of this paper.

REPRODUCIBILITY

Appendix C contains the hyperparameters of all fine-tuned models. Section 5.2 includes more details on the specific models and LLMs used. All LLM prompts used in ZEST are listed in Appendix B.

Section 5.1 explains the metrics used. We provide a detailed description of the preprocessing steps of the LEMONADE dataset in Section 3, and its statistics in Section 3 and Appendix A.

We are also attaching an anonymized version of our code for ZEST, and a sample of the LEMONADE dataset to this submission. We will publicly release the code and the full dataset upon publication.

REFERENCES

ACLED. Coding review process, September 2020. URL https://acleddata.com/acleddatanew/wp-content/uploads/2021/11/ACLED_Coding-Review-Process_v2_September-2020.pdf. Accessed: 2024-09-30.

ACLED. 2023 impact report, July 2023. URL https://acleddata.com/acleddatanew/wp-content/uploads/2024/07/2023-Impact-Report-Final-July-17.pdf. Accessed: 2024-09-28.

ACLED. Armed conflict location & event data project (acled) codebook. https://acleddata.com/knowledge-base/codebook/, 2023. Last updated: 27 September 2024.

Ujjwala Anantheswaran, Himanshu Gupta, Mihir Parmar, Kuntal Kumar Pal, and Chitta Baral. Edm3: Event detection as multi-task text generation, 2023.

Ashton Anderson, Dan Jurafsky, and Daniel A. McFarland. Towards a computational history of the ACL: 1980-2008. In Rafael E. Banchs (ed.), *Proceedings of the ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries*, pp. 13–21, Jeju Island, Korea, July 2012. Association for Computational Linguistics. URL https://aclanthology.org/W12-3202.

Theodora-Ismene Gizelis Andrea Ruggeri and Han Dorussen. Events data as bismarck's sausages? intercoder reliability, coders' selection, and data quality. *International Interactions*, 37(3):340–361, 2011. doi: 10.1080/03050629.2011.596028. URL https://doi.org/10.1080/03050629.2011.596028.

Edward E. Azar. The conflict and peace data bank (copdab) project. *Journal of Conflict Resolution*, 24(1):143–152, 1980. doi: 10.1177/002200278002400106. URL https://doi.org/10.1177/002200278002400106.

Femke Bakker, Ruben Van Heusden, and Maarten Marx. Timeline extraction from decision letters using ChatGPT. In Ali Hürriyetoğlu, Hristo Tanev, Surendrabikram Thapa, and Gökçe Uludoğan (eds.), *Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2024)*, pp. 24–31, St. Julians, Malta, March 2024. Association for Computational Linguistics. URL https://aclanthology.org/2024.case-1.3.

Ali Balali, Masoud Asadpour, and Seyed Hossein Jafari. Cofee: A comprehensive ontology for event extraction from text. *SSRN Electronic Journal*, 2022. ISSN 1556-5068. doi: 10.2139/ssrn.4117538. URL http://dx.doi.org/10.2139/ssrn.4117538.

Giovanni Campagna, Sina Semnani, Ryan Kearns, Lucas Jun Koba Sato, Silei Xu, and Monica Lam. A few-shot semantic parser for Wizard-of-Oz dialogues with the precise ThingTalk representation. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (eds.), *Findings of the Association for Computational Linguistics: ACL 2022*, pp. 4021–4034, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.findings-acl.317. URL https://aclanthology.org/2022.findings-acl.317.

Tommaso Caselli and Chu-Ren Huang. Sourcing the crowd for a few good ones: Event type detection. In Martin Kay and Christian Boitet (eds.), *Proceedings of COLING 2012: Posters*, pp. 1239–1248, Mumbai, India, December 2012. The COLING 2012 Organizing Committee. URL https://aclanthology.org/C12-2121.

Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. Event extraction via dynamic multi-pooling convolutional neural networks. In Chengqing Zong and Michael Strube (eds.), *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 167–176, Beijing, China, July 2015. Association for Computational Linguistics. doi: 10.3115/v1/P15-1017. URL https://aclanthology.org/P15-1017.

Erica Chenoweth and Orion A Lewis. Unpacking nonviolent campaigns: Introducing the navco 2.0 dataset. *Journal of Peace Research*, 50(3):415–423, 2013.

Jonathan H. Clark, Eunsol Choi, Michael Collins, Dan Garrette, Tom Kwiatkowski, Vitaly Nikolaev, and Jennimaria Palomaki. TyDi QA: A benchmark for information-seeking question answering in typologically diverse languages. *Transactions of the Association for Computational Linguistics*, 8:454–470, 2020. doi: 10.1162/tacl_a_00317. URL https://aclanthology.org/2020.tacl-1.30.

Samuel Colvin, Eric Jolibois, Hasan Ramezani, Adrian Garcia Badaracco, Terrence Dorsey, David Montague, Serge Matveenko, Marcelo Trylesinski, Sydney Runkle, David Hewitt, and Alex Hall. Pydantic, September 2024. URL https://docs.pydantic.dev/latest/.

Sergio Consoli, Peter Markov, Nikolaos I. Stilianakis, Lorenzo Bertolini, Antonio Puertas Gallardo, and Mario Ceresa. Epidemic information extraction for event-based surveillance using large language models. In Xin-She Yang, Simon Sherratt, Nilanjan Dey, and Amit Joshi (eds.), *Proceedings of Ninth International Congress on Information and Communication Technology*, pp. 241–252, Singapore, 2024. Springer Nature Singapore. ISBN 978-981-97-4581-4.

Ethan Bueno de Mesquita, C. Christine Fair, Jenna Jordan, Rasul Bakhsh Rais, and Jacob N. Shapiro. Measuring political violence in pakistan: Insights from the bfrs dataset. *Conflict Management and Peace Science*, 32(5):536–558, 2015. doi: 10.1177/0738894214542401. URL https://doi.org/10.1177/0738894214542401.

Alexandra DeLucia, Mark Dredze, and Anna L. Buczak. A multi-instance learning approach to civil unrest event detection on Twitter. In Ali Hürriyetoğlu, Hristo Tanev, Vanni Zavarella, Reyyan Yeniterzi, Erdem Yörük, and Milena Slavcheva (eds.), *Proceedings of the 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text*, pp. 18–33, Varna, Bulgaria, September 2023. INCOMA Ltd., Shoumen, Bulgaria. URL https://aclanthology.org/2023.case-1.3.

Xinya Du and Claire Cardie. Event extraction by answering (almost) natural questions. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 671–683, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.49. URL https://aclanthology.org/2020.emnlp-main.49.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Lauren Rantala-Yeary, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kambadur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Olivier Duchenne, Onur Çelebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Raparthy, Sheng Shen, Shengye Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten Sootla, Stephane Collot, Suchin Gururangan, Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, Tobias Speckbacher, Todor Mihaylov, Tong Xiao, Ujjwal Karn, Vedanuj Goswami, Vibhor Gupta, Vignesh Ramanathan, Viktor Kerkez, Vincent Gonguet, Virginie Do, Vish Vogeti, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaoqing Ellen Tan, Xinfeng Xie, Xuchao Jia, Xuewei Wang, Yaelle Goldschlag, Yashesh Gaur, Yasmine Babaei, Yi Wen, Yiwen Song, Yuchen Zhang, Yue Li, Yuning Mao, Zacharie Delpierre Coudert, Zheng Yan, Zhengxing Chen, Zoe Papakipos, Aaditya Singh, Aaron Grattafiori, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenberg, Alex Vaughan, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Franco, Aparajita Saraf, Arkabandhu Chowdhury, Ashley Gabriel, Ashwin Bharambe, Assaf Eisenman, Azadeh Yazdan, Beau James, Ben Maurer, Benjamin Leonhardi, Bernie Huang, Beth Loyd, Beto De Paola, Bhargavi Paranjape, Bing Liu, Bo Wu, Boyu Ni, Braden Hancock, Bram Wasti, Brandon Spence, Brani Stojkovic, Brian Gamido, Britt Montalvo, Carl Parker, Carly Burton, Catalina Mejia, Changhan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, Danny Wyatt, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, Dingkang Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Montgomery, Eleonora Presani, Emily Hahn, Emily Wood, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Firat Ozgenel, Francesco Caggioni, Francisco Guzmán, Frank Kanayet, Frank Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Govind Thattai, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Ibrahim Damlaj, Igor Molybog, Igor Tufanov, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, Japhet Asher, Jean-Baptiste Gaya, Jeff Marcus, Jeff Tang, Jennifer Chan, Jenny Zhen, Jeremy Reizenstein, Jeremy Teboul, Jessica Zhong, Jian Jin, Jingyi Yang, Joe Cummings, Jon Carvill, Jon Shepard, Jonathan McPhie, Jonathan Torres, Josh Ginsburg, Junjie Wang, Kai Wu, Kam Hou U, Karan Saxena, Karthik Prasad, Kartikay Khandelwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kun Huang, Kunal Chawla, Kushal Lakhotia, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabsa, Manav Avalani, Manish Bhatt, Maria Tsimpoukelli, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov, Maya Lathi, Meghan Keneally, Michael L. Seltzer, Michal Valko, Michelle Restrepo, Mihir Patel, Mik Vyatskov, Mikayel Samvelyan, Mike Clark, Mike Macey, Mike Wang, Miquel Jubert Hermoso, Mo Metanat, Mohammad Rastegari, Munish Bansal, Nandhini Santhanam, Natascha Parks, Natasha White, Navyata Bawa, Nayan Singhal, Nick Egebo, Nicolas Usunier, Nikolay Pavlovich Laptev, Ning Dong, Ning Zhang, Norman Cheng, Oleg Chernoguz, Olivia Hart, Omkar Salpekar, Ozlem Kalinli, Parkin Kent, Parth Parekh, Paul Saab, Pavan Balaji, Pedro Rittner, Philip Bontrager, Pierre Roux, Piotr Dollar, Polina Zvyagina, Prashant Ratanchandani, Pritish Yuvraj, Qian Liang, Rachad Alao, Rachel Rodriguez, Rafi Ayub, Raghotham Murthy, Raghu Nayani, Rahul Mitra, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Rohan Maheswari, Russ Howes, Ruty Rinott, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, Shiva Shankar, Shuqiang Zhang, Shuqiang Zhang, Sinong Wang, Sneha Agarwal, Soji Sajuyigbe, Soumith Chintala, Stephanie Max, Stephen Chen, Steve Kehoe, Steve Satterfield, Sudarshan Govindaprasad, Sumit Gupta, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Kohler, Thomas Robinson, Tianhe Li, Tianjun Zhang, Tim Matthews, Timothy Chou, Tzook Shaked, Varun Vontimitta, Victoria Ajayi, Victoria Montanez, Vijai Mohan, Vinay Satish Kumar, Vishal Mangla, Vítor Albiero, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaofang Wang, Xiaojian Wu, Xiaolan Wang, Xide Xia, Xilun Wu, Xinbo Gao, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yuchen Hao, Yundi Qian, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, and Zhiwei Zhao. The llama 3 herd of models, 2024. URL `https://arxiv.org/abs/2407.21783`.

Fırat Duruşan, Ali Hürriyetoğlu, Erdem Yörük, Osman Mutlu, Çağrı Yoltar, Burak Gürel, and Alvaro Comin. Global contentious politics database (glocon) annotation manuals, 2022.

Sanghamitra Dutta, Liang Ma, Tanay Kumar Saha, Di Liu, Joel Tetreault, and Alejandro Jaimes. GTN-ED: Event detection using graph transformer networks. In Alexander Panchenko, Fragkiskos D. Malliaros, Varvara Logacheva, Abhik Jana, Dmitry Ustalov, and Peter Jansen (eds.), *Proceedings of the Fifteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-15)*, pp. 132–137, Mexico City, Mexico, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.textgraphs-1.13. URL `https://aclanthology.org/2021.textgraphs-1.13`.

Deborah J. Gerner, Philip A. Schrodt, and Ömür Yilmaz. *Conflict and mediation event observations (CAMEO): An event data framework for a post-cold war world*, pp. 287–304. Routledge Taylor & Francis Group, December 2008. ISBN 0203885139. doi: 10.4324/9780203885130.

Ralph Grishman and Beth Sundheim. Message Understanding Conference- 6: A brief history. In *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*, 1996. URL `https://aclanthology.org/C96-1079`.

Johan Dittrich Hallberg. Prio conflict site 1989–2008: A geo-referenced dataset on armed conflict. *Conflict Management and Peace Science*, 29(2):219–232, 2012. doi: 10.1177/0738894211433168. URL `https://doi.org/10.1177/0738894211433168`.

Andrew Halterman, Benjamin E Bagozzi, Andreas Beger, Phil Schrodt, and Grace Scraborough. Plover and polecat: A new political event ontology and dataset, Apr 2023a. URL `osf.io/preprints/socarxiv/rm5dw`.

Andrew Halterman, Philip A. Schrodt, Andreas Beger, Benjamin E. Bagozzi, and Grace I. Scarborough. Creating custom event data without dictionaries: A bag-of-tricks, 2023b.

Luheng He, Mike Lewis, and Luke Zettlemoyer. Question-answer driven semantic role labeling: Using natural language to annotate natural language. In Lluís Màrquez, Chris Callison-Burch, and Jian Su (eds.), *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 643–653, Lisbon, Portugal, September 2015. Association for Computational Linguistics. doi: 10.18653/v1/D15-1076. URL `https://aclanthology.org/D15-1076`.

I-Hung Hsu, Kuan-Hao Huang, Elizabeth Boschee, Scott Miller, Prem Natarajan, Kai-Wei Chang, and Nanyun Peng. DEGREE: A data-efficient generation-based event extraction model. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (eds.), *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1890–1908, Seattle, United States, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-main.138. URL `https://aclanthology.org/2022.naacl-main.138`.

Yibo Hu, MohammadSaleh Hosseini, Erick Skorupa Parolin, Javier Osorio, Latifur Khan, Patrick Brandt, and Vito D'Orazio. ConfliBERT: A pre-trained language model for political conflict and violence. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (eds.), *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 5469–5482, Seattle, United States, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-main.400. URL `https://aclanthology.org/2022.naacl-main.400`.

Kuan-Hao Huang, I-Hung Hsu, Tanmay Parekh, Zhiyu Xie, Zixuan Zhang, Prem Natarajan, Kai-Wei Chang, Nanyun Peng, and Heng Ji. TextEE: Benchmark, reevaluation, reflections, and future challenges in event extraction. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Findings of the Association for Computational Linguistics ACL 2024*, pp. 12804–12825, Bangkok, Thailand and virtual meeting, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-acl.760. URL https://aclanthology.org/2024.findings-acl.760.

Ali Hürriyetoğlu (ed.). *Proceedings of the 4th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2021)*, Online, August 2021. Association for Computational Linguistics. URL https://aclanthology.org/2021.case-1.0.

Ali Hürriyetoğlu, Osman Mutlu, Fırat Duruşan, Onur Uca, Alaeddin Gürel, Benjamin J. Radford, Yaoyao Dai, Hansi Hettiarachchi, Niklas Stoehr, Tadashi Nomoto, Milena Slavcheva, Francielle Vargas, Aaqib Javid, Fatih Beyhan, and Erdem Yörük. Extended multilingual protest news detection - shared task 1, CASE 2021 and 2022. In Ali Hürriyetoğlu, Hristo Tanev, Vanni Zavarella, and Erdem Yörük (eds.), *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE)*, pp. 223–228, Abu Dhabi, United Arab Emirates (Hybrid), December 2022a. Association for Computational Linguistics. doi: 10.18653/v1/2022.case-1.31. URL https://aclanthology.org/2022.case-1.31.

Ali Hürriyetoğlu, Hristo Tanev, Vanni Zavarella, and Erdem Yörük (eds.). *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE)*, Abu Dhabi, United Arab Emirates (Hybrid), December 2022b. Association for Computational Linguistics. URL https://aclanthology.org/2022.case-1.0.

Ali Hürriyetoğlu, Hristo Tanev, Vanni Zavarella, Reyyan Yeniterzi, Erdem Yörük, and Milena Slavcheva (eds.). *Proceedings of the 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text*, Varna, Bulgaria, September 2023. INCOMA Ltd., Shoumen, Bulgaria. URL https://aclanthology.org/2023.case-1.0.

Ali Hürriyetoğlu, Hristo Tanev, Surendrabikram Thapa, and Gökçe Uludoğan (eds.). *Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2024)*, St. Julians, Malta, March 2024a. Association for Computational Linguistics. URL https://aclanthology.org/2024.case-1.0.

Ali Hürriyetoğlu, Surendrabikram Thapa, Gökçe Uludoğan, Somaiyeh Dehghan, and Hristo Tanev. A concise report of the 7th workshop on challenges and applications of automated extraction of socio-political events from text. In Ali Hürriyetoğlu, Hristo Tanev, Surendrabikram Thapa, and Gökçe Uludoğan (eds.), *Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2024)*, pp. 248–255, St. Julians, Malta, March 2024b. Association for Computational Linguistics. URL https://aclanthology.org/2024.case-1.34.

Heng Ji and Ralph Grishman. Refining event extraction through cross-document inference. In Johanna D. Moore, Simone Teufel, James Allan, and Sadaoki Furui (eds.), *Proceedings of ACL-08: HLT*, pp. 254–262, Columbus, Ohio, June 2008. Association for Computational Linguistics. URL https://aclanthology.org/P08-1030.

J. D. Kim, T. Ohta, Y. Tateisi, and J. Tsujii. Genia corpus—a semantically annotated corpus for bio-textmining. *Bioinformatics*, 19(suppl 1):i180–i182, 07 2003. ISSN 1367-4803. doi: 10.1093/bioinformatics/btg1023. URL https://doi.org/10.1093/bioinformatics/btg1023.

Hanspeter Kriesi, Bruno Wüest, Jasmine Lorenzini, Peter Makarov, Matthias Enggist, Klaus Rothenhäusler, Thomas Kurer, Silja Häusermann, and Altiparmakis Patrice Wangen. Poldem–protest event dataset 30, 2019.

Viet Dac Lai. Event extraction: A survey. *arXiv preprint arXiv:2210.03419*, 2022.

Viet Dac Lai, Tuan Ngo Nguyen, and Thien Huu Nguyen. Event detection: Gate diversity and syntactic importance scores for graph convolution neural networks. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 5405–5411, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.435. URL https://aclanthology.org/2020.emnlp-main.435.

Kalev Leetaru and Philip A. Schrodt. Gdelt: Global data on events, location, and tone. *ISA Annual Convention*, 2013. URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.686.6605`.

Manling Li, Ying Lin, Joseph Hoover, Spencer Whitehead, Clare Voss, Morteza Dehghani, and Heng Ji. Multilingual entity, relation, event and human value extraction. In Waleed Ammar, Annie Louis, and Nasrin Mostafazadeh (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pp. 110–115, Minneapolis, Minnesota, June 2019a. Association for Computational Linguistics. doi: 10.18653/v1/N19-4019. URL `https://aclanthology.org/N19-4019`.

Manling Li, Alireza Zareian, Ying Lin, Xiaoman Pan, Spencer Whitehead, Brian Chen, Bo Wu, Heng Ji, Shih-Fu Chang, Clare Voss, Daniel Napierski, and Marjorie Freedman. GAIA: A fine-grained multimedia knowledge extraction system. In Asli Celikyilmaz and Tsung-Hsien Wen (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 77–86, Online, July 2020a. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-demos.11. URL `https://aclanthology.org/2020.acl-demos.11`.

Manling Li, Qi Zeng, Ying Lin, Kyunghyun Cho, Heng Ji, Jonathan May, Nathanael Chambers, and Clare Voss. Connecting the dots: Event graph schema induction with path language modeling. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 684–695, Online, November 2020b. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.50. URL `https://aclanthology.org/2020.emnlp-main.50`.

Manling Li, Sha Li, Zhenhailong Wang, Lifu Huang, Kyunghyun Cho, Heng Ji, Jiawei Han, and Clare Voss. The future is not one-dimensional: Complex event schema induction by graph modeling for event prediction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 5203–5215, 2021a.

Qian Li, Jianxin Li, Jiawei Sheng, Shiyao Cui, Jia Wu, Yiming Hei, Hao Peng, Shu Guo, Lihong Wang, Amin Beheshti, et al. A survey on deep learning event extraction: Approaches and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

Sha Li, Heng Ji, and Jiawei Han. Document-level event argument extraction by conditional generation. In Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (eds.), *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 894–908, Online, June 2021b. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.69. URL `https://aclanthology.org/2021.naacl-main.69`.

Xiaoya Li, Fan Yin, Zijun Sun, Xiayu Li, Arianna Yuan, Duo Chai, Mingxin Zhou, and Jiwei Li. Entity-relation extraction as multi-turn question answering. In Anna Korhonen, David Traum, and Lluís Màrquez (eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 1340–1350, Florence, Italy, July 2019b. Association for Computational Linguistics. doi: 10.18653/v1/P19-1129. URL `https://aclanthology.org/P19-1129`.

Shasha Liao and Ralph Grishman. Acquiring topic features to improve event extraction: in pre-selected and balanced collections. In Ruslan Mitkov and Galia Angelova (eds.), *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*, pp. 9–16, Hissar, Bulgaria, September 2011. Association for Computational Linguistics. URL `https://aclanthology.org/R11-1002`.

Jiangwei Liu, Liangyu Min, and Xiaohong Huang. An overview of event extraction and its applications, 2021.

Xiao Liu, Zhunchen Luo, and Heyan Huang. Jointly multiple events extraction via attention-based graph information aggregation. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language*

Processing, pp. 1247–1256, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1156. URL https://aclanthology.org/D18-1156.

Meagan Loerakker, Laurens Müter, and Marijn Schraagen. Fine-tuning language models on Dutch protest event tweets. In Ali Hürriyetoğlu, Hristo Tanev, Surendrabikram Thapa, and Gökçe Uludoğan (eds.), *Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2024)*, pp. 6–23, St. Julians, Malta, March 2024. Association for Computational Linguistics. URL https://aclanthology.org/2024.case-1.2.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2017.

Yaojie Lu, Hongyu Lin, Jin Xu, Xianpei Han, Jialong Tang, Annan Li, Le Sun, Meng Liao, and Shaoyi Chen. Text2Event: Controllable sequence-to-structure generation for end-to-end event extraction. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 2795–2806, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.217. URL https://aclanthology.org/2021.acl-long.217.

Yinquan Lu, Wenhao Zhu, Lei Li, Yu Qiao, and Fei Yuan. Llamax: Scaling linguistic horizons of llm by enhancing translation capabilities beyond 100 languages, 2024.

Kevin Lybarger, Mari Ostendorf, and Meliha Yetisgen. Annotating social determinants of health using active learning, and characterizing determinants using neural event extraction. *Journal of Biomedical Informatics*, 113:103631, Jan 2021. ISSN 1532-0464. doi: 10.1016/j.jbi.2020.103631. URL https://doi.org/10.1016/j.jbi.2020.103631. Epub 2020 Dec 5.

Ayush Maheshwari, Hrishikesh Patel, Nandan Rathod, Ritesh Kumar, Ganesh Ramakrishnan, and Pushpak Bhattacharyya. Tale of tails using rule augmented sequence labeling for event extraction, 2019.

Alex Mallen, Akari Asai, Victor Zhong, Rajarshi Das, Daniel Khashabi, and Hannaneh Hajishirzi. When not to trust language models: Investigating effectiveness of parametric and non-parametric memories. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 9802–9822, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.546. URL https://aclanthology.org/2023.acl-long.546.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, USA, 2008. ISBN 0521865719.

Charles A. McClelland. World event/interaction survey. 1978. URL https://api.semanticscholar.org/CorpusID:154928040.

Sneha Mehta, Huzefa Rangwala, and Naren Ramakrishnan. Improving zero-shot event extraction via sentence simplification. In Ali Hürriyetoğlu, Hristo Tanev, Vanni Zavarella, and Erdem Yörük (eds.), *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE)*, pp. 32–43, Abu Dhabi, United Arab Emirates (Hybrid), December 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.case-1.5. URL https://aclanthology.org/2022.case-1.5.

Julian Michael, Gabriel Stanovsky, Luheng He, Ido Dagan, and Luke Zettlemoyer. Crowdsourcing question-answer meaning representations. In Marilyn Walker, Heng Ji, and Amanda Stent (eds.), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pp. 560–568, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-2089. URL https://aclanthology.org/N18-2089.

Bonan Min, Benjamin Rozonoyer, Haoling Qiu, Alexander Zamanian, Nianwen Xue, and Jessica MacBride. ExcavatorCovid: Extracting events and relations from text corpora for temporal and causal analysis for COVID-19. In Heike Adel and Shuming Shi (eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 63–71, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.emnlp-demo.8. URL https://aclanthology.org/2021.emnlp-demo.8.

Mehrad Moradshahi, Giovanni Campagna, Sina Semnani, Silei Xu, and Monica Lam. Localizing open-ontology QA semantic parsers in a day using machine translation. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 5970–5983, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.481. URL https://aclanthology.org/2020.emnlp-main.481.

Aida Mostafazadeh Davani, Leigh Yeh, Mohammad Atari, Brendan Kennedy, Gwenyth Portillo Wightman, Elaine Gonzalez, Natalie Delong, Rhea Bhatia, Arineh Mirinjian, Xiang Ren, and Morteza Dehghani. Reporting the unreported: Event extraction for analyzing the local representation of hate crimes. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 5753–5757, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1580. URL https://aclanthology.org/D19-1580.

Osman Mutlu and Ali Hürriyetoğlu. Negative documents are positive: Improving event extraction performance using overlooked negative data. In Ali Hürriyetoğlu, Hristo Tanev, Vanni Zavarella, Reyyan Yeniterzi, Erdem Yörük, and Milena Slavcheva (eds.), *Proceedings of the 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text*, pp. 124–135, Varna, Bulgaria, September 2023. INCOMA Ltd., Shoumen, Bulgaria. URL https://aclanthology.org/2023.case-1.17.

Minh Van Nguyen, Bonan Min, Franck Dernoncourt, and Thien Nguyen. Joint extraction of entities, relations, and events via modeling inter-instance and inter-label dependencies. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (eds.), *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4363–4374, Seattle, United States, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-main.324. URL https://aclanthology.org/2022.naacl-main.324.

Thi-Nhung Nguyen, Bang Tien Tran, Trong-Nghia Luu, Thien Huu Nguyen, and Kiem-Hieu Nguyen. BKEE: Pioneering event extraction in the Vietnamese language. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue (eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 2421–2427, Torino, Italia, May 2024. ELRA and ICCL. URL https://aclanthology.org/2024.lrec-main.217.

OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org, 2017.

Tanmay Parekh, Anh Mac, Jiarui Yu, Yuxuan Dong, Syed Shahriar, Bonnie Liu, Eric Yang, Kuan-Hao Huang, Wei Wang, Nanyun Peng, and Kai-Wei Chang. Event detection from social media for epidemic prediction. In Kevin Duh, Helena Gomez, and Steven Bethard (eds.), *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 5758–5783, Mexico City, Mexico, June 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.naacl-long.322. URL https://aclanthology.org/2024.naacl-long.322.

Kevin Pei, Ishan Jindal, and Kevin Chang. Abstractive open information extraction. In Houda Bouamor, Juan Pino, and Kalika Bali (eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 6146–6158, Singapore, December 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.emnlp-main.376. URL https://aclanthology.org/2023.emnlp-main.376.

Amir Pouran Ben Veyseh, Javid Ebrahimi, Franck Dernoncourt, and Thien Nguyen. MEE: A novel multilingual event extraction dataset. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang (eds.), *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 9603–9613, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.emnlp-main.652. URL https://aclanthology.org/2022.emnlp-main.652.

Yunjia Qi, Hao Peng, Xiaozhi Wang, Bin Xu, Lei Hou, and Juanzi Li. Adelie: Aligning large language models on information extraction, 2024.

Maxim Rabinovich, Mitchell Stern, and Dan Klein. Abstract syntax networks for code generation and semantic parsing. In Regina Barzilay and Min-Yen Kan (eds.), *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1139–1149, Vancouver, Canada, July 2017. Association for Computational Linguistics. doi: 10.18653/v1/P17-1105. URL https://aclanthology.org/P17-1105.

Dragomir R. Radev, Eduard Hovy, and Kathleen McKeown. Introduction to the special issue on summarization. *Computational Linguistics*, 28(4):399–408, 2002. doi: 10.1162/089120102762671927. URL https://aclanthology.org/J02-4001.

Benjamin J. Radford. Multitask models for supervised protests detection in texts, 2020.

Ria Raj, Kajsa Andreasson, Tobias Norlund, Richard Johansson, and Aron Lagerberg. Cross-modal transfer between vision and language for protest detection. In Ali Hürriyetoğlu, Hristo Tanev, Vanni Zavarella, and Erdem Yörük (eds.), *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE)*, pp. 56–60, Abu Dhabi, United Arab Emirates (Hybrid), December 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.case-1.8. URL https://aclanthology.org/2022.case-1.8.

Clionadh Raleigh, Rew Linke, Håvard Hegre, and Joakim Karlsen. Introducing acled: An armed conflict location and event dataset. *Journal of peace research*, 47(5):651–660, 2010.

Clionadh Raleigh, Roudabeh Kishi, and Andrew Linke. Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices. *Humanities and Social Sciences Communications*, 10(1):74, 2023. ISSN 2662-9992. doi: 10.1057/s41599-023-01559-4. URL https://doi.org/10.1057/s41599-023-01559-4.

Jeff Rasley, Samyam Rajbhandari, Olatunji Ruwase, and Yuxiong He. Deepspeed: System optimizations enable training deep learning models with over 100 billion parameters. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '20, pp. 3505–3506, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450379984. doi: 10.1145/3394486.3406703. URL https://doi.org/10.1145/3394486.3406703.

Revanth Gangi Reddy, Yi R Fung, Qi Zeng, Manling Li, Ziqi Wang, Paul Sullivan, and Heng Ji. Smartbook: Ai-assisted situation report generation. *arXiv preprint arXiv:2303.14337*, 2023.

Yubing Ren, Yanan Cao, Hao Li, Yingjie Li, Zixuan ZM Ma, Fang Fang, Ping Guo, and Wei Ma. DEIE: Benchmarking document-level event information extraction with a large-scale Chinese news dataset. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue (eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 4592–4604, Torino, Italia, May 2024. ELRA and ICCL. URL https://aclanthology.org/2024.lrec-main.410.

Oscar Sainz, Iker García-Ferrero, Rodrigo Agerri, Oier Lopez de Lacalle, German Rigau, and Eneko Agirre. GoLLIE: Annotation guidelines improve zero-shot information-extraction. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=Y3wpuxd7u9.

Sam Jones. New expansion brings ACLED to full global coverage, 2022. URL https://acleddata.com/2022/02/02/new-expansion-brings-acled-to-full-global-coverage/. Accessed: 2024-09-28.

Taneeya Satyapanich, Francis Ferraro, and Tim Finin. Casie: Extracting cybersecurity event information from text. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):8749–8757, Apr. 2020. doi: 10.1609/aaai.v34i05.6401. URL https://ojs.aaai.org/index.php/AAAI/article/view/6401.

Ge Shi, Yunyue Su, Yongliang Ma, and Ming Zhou. A hybrid detection and generation framework with separate encoders for event extraction. In Andreas Vlachos and Isabelle Augenstein (eds.), *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 3163–3180, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.eacl-main.231. URL https://aclanthology.org/2023.eacl-main.231.

Milena Slavcheva, Hristo Tanev, and Onur Uca. On the road to a protest event ontology for Bulgarian: Conceptual structures and representation design. In Ali Hürriyetoğlu, Hristo Tanev, Vanni Zavarella, Reyyan Yeniterzi, Erdem Yörük, and Milena Slavcheva (eds.), *Proceedings of the 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text*, pp. 92–100, Varna, Bulgaria, September 2023. INCOMA Ltd., Shoumen, Bulgaria. URL https://aclanthology.org/2023.case-1.13.

Hristo Tanev. Leveraging approximate pattern matching with BERT for event detection. In Ali Hürriyetoğlu, Hristo Tanev, Surendrabikram Thapa, and Gökçe Uludoğan (eds.), *Proceedings of the 7th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE 2024)*, pp. 32–39, St. Julians, Malta, March 2024. Association for Computational Linguistics. URL https://aclanthology.org/2024.case-1.4.

Hristo Tanev and Bertrand De Longueville. Where "where" matters : Event location disambiguation with a BERT language model. In Ali Hürriyetoğlu, Hristo Tanev, Vanni Zavarella, Reyyan Yeniterzi, Erdem Yörük, and Milena Slavcheva (eds.), *Proceedings of the 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text*, pp. 11–17, Varna, Bulgaria, September 2023. INCOMA Ltd., Shoumen, Bulgaria. URL https://aclanthology.org/2023.case-1.2.

MeiHan Tong, Bin Xu, Shuai Wang, Meihuan Han, Yixin Cao, Jiangqi Zhu, Siyu Chen, Lei Hou, and Juanzi Li. DocEE: A large-scale and fine-grained benchmark for document-level event extraction. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz (eds.), *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 3970–3982, Seattle, United States, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.naacl-main.291. URL https://aclanthology.org/2022.naacl-main.291.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. Llama 2: Open foundation and fine-tuned chat models, 2023.

Jennifer Tracey, Ann Bies, Jeremy Getman, Kira Griffitt, and Stephanie Strassel. A study in contradiction: Data and annotation for AIDA focusing on informational conflict in Russia-Ukraine relations. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis (eds.), *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 1831–1838, Marseille, France, June 2022. European Language Resources Association. URL https://aclanthology.org/2022.lrec-1.195.

Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda. Ace 2005 multilingual training corpus ldc2006t06, 2006.

Xingyao Wang, Sha Li, and Heng Ji. Code4Struct: Code generation for few-shot event structure prediction. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 3640–3663, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.202. URL https://aclanthology.org/2023.acl-long.202.

Chih-Hsuan Wei, Yifan Peng, Robert Leaman, Allan Peter Davis, Carolyn J Mattingly, Jiao Li, Thomas C Wiegers, and Zhiyong Lu. Assessing the state of the art in biomedical relation extraction: overview of the biocreative v chemical-disease relation (cdr) task. *Database (Oxford)*, 2016: baw032, 2016. doi: 10.1093/database/baw032. URL http://www.biocreative.org/tasks/biocreative-v/track-3-cdr/. Published by Oxford University Press 2016. This work is written by US Government employees and is in the public domain in the US.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models, 2023. URL https://arxiv.org/abs/2201.11903.

Nils B Weidmann and Espen Geelmuyden Rød. *The Internet and political protest in autocracies*. Oxford Studies in Digital Poli, 2019.

Haoyang Wen, Ying Lin, Tuan Lai, Xiaoman Pan, Sha Li, Xudong Lin, Ben Zhou, Manling Li, Haoyu Wang, Hongming Zhang, Xiaodong Yu, Alexander Dong, Zhenhailong Wang, Yi Fung, Piyush Mishra, Qing Lyu, Dídac Surís, Brian Chen, Susan Windisch Brown, Martha Palmer, Chris Callison-Burch, Carl Vondrick, Jiawei Han, Dan Roth, Shih-Fu Chang, and Heng Ji. RESIN: A dockerized schema-guided cross-document cross-lingual cross-media information extraction and event tracking system. In Avi Sil and Xi Victoria Lin (eds.), *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstrations*, pp. 133–143, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-demos.16. URL https://aclanthology.org/2021.naacl-demos.16.

Brandon T. Willard and Rémi Louf. Efficient guided generation for large language models, 2023.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. Huggingface's transformers: State-of-the-art natural language processing, 2019.

Derong Xu, Wei Chen, Wenjun Peng, Chao Zhang, Tong Xu, Xiangyu Zhao, Xian Wu, Yefeng Zheng, Yang Wang, and Enhong Chen. Large language models for generative information extraction: A survey, 2023.

Sen Yang, Dawei Feng, Linbo Qiao, Zhigang Kan, and Dongsheng Li. Exploring pre-trained language models for event extraction and generation. In Anna Korhonen, David Traum, and Lluís Màrquez (eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 5284–5294, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1522. URL https://aclanthology.org/P19-1522.

Vanni Zavarella, Dilek Küçük, Hristo Tanev, and Ali Hürriyetoğlu. Event extraction for Balkan languages. In Shuly Wintner, Marko Tadić, and Bogdan Babych (eds.), *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational*

Linguistics, pp. 65–68, Gothenburg, Sweden, April 2014. Association for Computational Linguistics. doi: 10.3115/v1/E14-2017. URL `https://aclanthology.org/E14-2017`.

Vanni Zavarella, Hristo Tanev, Ali Hürriyetoğlu, Peratham Wiriyathammabhum, and Bertrand De Longueville. Tracking COVID-19 protest events in the United States. shared task 2: Event database replication, CASE 2022. In Ali Hürriyetoğlu, Hristo Tanev, Vanni Zavarella, and Erdem Yörük (eds.), *Proceedings of the 5th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE)*, pp. 209–216, Abu Dhabi, United Arab Emirates (Hybrid), December 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.case-1.29. URL `https://aclanthology.org/2022.case-1.29`.

Hongming Zhang, Xin Liu, Haojie Pan, Yangqiu Song, and Cane Wing-Ki Leung. Aser: A large-scale eventuality knowledge graph. In *Proceedings of The Web Conference 2020*, WWW '20, pp. 201–211, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450370233. doi: 10.1145/3366423.3380107. URL `https://doi.org/10.1145/3366423.3380107`.

Zixuan Zhang and Heng Ji. Abstract Meaning Representation guided graph encoding and decoding for joint information extraction. In Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tur, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (eds.), *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 39–49, Online, June 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.naacl-main.4. URL `https://aclanthology.org/2021.naacl-main.4`.

Shun Zheng, Wei Cao, Wei Xu, and Jiang Bian. Doc2EDAG: An end-to-end document-level framework for Chinese financial event extraction. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 337–346, Hong Kong, China, November 2019. Association for Computational Linguistics. doi: 10.18653/v1/D19-1032. URL `https://aclanthology.org/D19-1032`.

Han Zhou, Hongpeng Yin, Hengyi Zheng, and Yanxia Li. A survey on multi-modal social event detection. *Knowledge-Based Systems*, 195:105695, 2020.

Mengna Zhu, Zijie Xu, Kaisheng Zeng, Kaiming Xiao, Mao Wang, Wenjun Ke, and Hongbin Huang. CMNEE:a large-scale document-level event extraction dataset based on open-source Chinese military news. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue (eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 3367–3379, Torino, Italia, May 2024a. ELRA and ICCL. URL `https://aclanthology.org/2024.lrec-main.299`.

Mengna Zhu, Zijie Xu, Kaisheng Zeng, Kaiming Xiao, Mao Wang, Wenjun Ke, and Hongbin Huang. Cmnee: A large-scale document-level event extraction dataset based on open-source chinese military news. *arXiv preprint arXiv:2404.12242*, 2024b.

Yuri M Zhukov, Christian Davenport, and Nadiya Kostyuk. Introducing xsub: A new portal for cross-national data on subnational violence. *Journal of Peace Research*, 56(4):604–614, 2019. doi: 10.1177/0022343319836697. URL `https://doi.org/10.1177/0022343319836697`.

Arkaitz Zubiaga, Damiano Spina, Raquel Martínez, and Víctor Fresno. Real-time classification of twitter trends. *Journal of the Association for Information Science and Technology*, 66(3):462–473, May 2014. ISSN 2330-1643. doi: 10.1002/asi.23186. URL `http://dx.doi.org/10.1002/asi.23186`.

A LEMONADE STATISTICS

Tables 4 and Figure 2 show the distribution of event types and country-level locations of events in LEMONADE respectively.

Table 4: The number of event types in all splits of LEMONADE. While imbalanced, the distribution of event types tracks that of the real world. For example, by far the most common among these events are peaceful protests.

Event Type	Count
GovernmentRegainsTerritory	6
NonStateActorOvertakesTerritory	55
ArmedClash	2775
ExcessiveForceAgainstProtestors	30
ProtestWithIntervention	993
PeacefulProtest	24805
ViolentDemonstration	910
MobViolence	2015
AirOrDroneStrike	1218
SuicideBomb	4
ShellingOrArtilleryOrMissileAttack	1161
RemoteExplosiveOrLandmineOrIED	480
Grenade	93
SexualViolence	54
Attack	3231
AbductionOrForcedDisappearance	304
Agreement	68
Arrest	631
ChangeToArmedGroup	362
DisruptedWeaponsUse	641
BaseEstablished	12
LootingOrPropertyDestruction	780
NonViolentTransferOfTerritory	19
OtherStrategicDevelopment	500

(a) Train set (b) Validation set (c) Test set

Figure 2: Distribution of event locations in LEMONADE. Note that the dataset includes more specific locations, but here we only plot the country level. In addition to being linguistically diverse, LEMONADE is also geographically diverse. The distribution of the train set is skewed towards India, because it only contains English events.

B ALL PROMPTS FOR ZEST

Here we provide the prompts used in ZEST. Some prompts are edited for brevity. The full text of prompts can be obtained from our code. The syntax used is the Jinja2 template language, which supports Python-like loops (`{% for %}{% endfor %}`), conditions (`{% if %}{% endif %}`), variables (`{{ var }}`) and comments (`#`).

```
# instruction
You are tasked with determining the best matching Event types for a given
     news article. You will be provided with annotation guidelines and a
    news article to analyze. Your goal is to identify the most relevant
    event types and rank them in order of their match to the article
    content.

# input
Here is the news article you need to analyze:
{{ article }}

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

{% for ed in event_definitions.items() %}
[{{ loop.index }}] "{{ ed[0] }}": {{ ed[1] }}

{% endfor %}


1. For each event type, determine how well it matches the article content
    . Consider the following factors:
  - How closely the event description aligns with the main focus of the
    article
  - The presence of key actors or entities mentioned in the event type
    description
  - The occurrence of specific actions or outcomes associated with the
    event type

2. Rank the event types based on their relevance to the article content.
    Only include event types that have a meaningful connection to the
    article.

3. Output your results using the following format:
   - List the relevant event types in descending order of match quality
   - Use the ">" symbol to separate the event types

Your output should look like this:

[Explain your reasoning for the event types you decide to include, and
    their order]

event_type_1 > event_type_2 > ...

Remember to exclude any event types that are not relevant to the article
    content. Provide only the ranked list of event types in your final
    answer.
```

Table 5: Prompt for event type detection (ED).

```
# 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", "Labour Group", etc.

# input
News article:
{{ article }}

The event you should focus on is the {{ event }} 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:

{% for entity in potential_entities %}
[{{ loop.index }}] {{ entity }}
{% endfor %}


Provide your answer listing one entity name per line:
entity name 1
entity name 2
...
```

Table 6: Prompt for the first stage of Entity Detection and Linking (EDL).

```
# instruction
In this task, an "entity" refers to an individual, group, or entity
    involved in the event described in the news article. entities can
    include:

1. State forces
2. Rebels
3. Militias
4. Identity groups
5. Demonstrators
6. Civilians
7. External or other forces

Most entities in political violence events are organized armed groups
    with a political purpose. They use violence for political means and
    are typically named entities. However, entities can also include
    unorganized groups like rioters, as well as protestors and civilians.

Your task is to find supporting evidence for each of the specified
    entities in the given article.

# input
Follow these steps carefully:

1. First, you will be provided with the full text of the news article:

{{ article }}

2. Next, you will be given a list of entities involved with the {{
    event_type }} event to search for:

{% for e in entities %}
   {{ e }}
{% endfor %}

3. Identify all supporting evidence of each given entity. These could be
    spans involving:
   - The exact entity name or variations of its name
   - Descriptive phrases that identify the entity
   - Phrases that could be used to infer the involvement of the entity

4. If there are multiple evidence for the involvement of an entity,
    output all of them.

5. For each evidence you find for an entity, provide your answer in the
    provided structure.
   Notes:
   - Include the original entity name in the `entity_original` field to
    denote which entities the evidence is for.
   - The character index starts at 0 for the first character of the
    article.
   - If there are multiple evidences for an entity, provide multiple `
    entitiespan`s for it.
   - If no evidence is found for an entity, respond with a mostly empty `
    entitiespan` and only fill the `explanation` field.

Remember to be precise in your span detection and provide clear
    explanations for each evidence span.
```

Table 7: Prompt for the second stage of Entity Detection and Linking (EDL).

```
# instruction
An "entity" refers to an individual, group, or entity involved in an event. Most entities in political violence events are organized armed groups with a political purpose. They use violence for political means and are typically named entities. However, entities can also include unorganized groups like Rioters, Protestors and Civilians.
An entity can be a generic term like "Students" or "Protestors", or a specific political group, militia, or armed group
Never use an individual's name as an entity. If a politician is mentioned, use the name of the political party or group they belong to.
Sometimes, a specific entity is accompanied by a generic entity. For example, a political party leading a protest should be counted as two actors: the political party, and "Protestors".
You will be given a news article, an event extracted from it, and a list of actors. Your task is to assign each entity to the correct event argument based on the information provided in the news article.

# input
First, carefully read the following news article:
{{ article }}

Now, consider the following event extracted from the article:
{{ event }}

Here is the list of actors to be assigned to event arguments:
{% for e in entities %}
- {{ e }}
{% endfor %}

You need to assign each actor to one of the following event arguments. Do not modify any other part of the event.
{% for field in actor_fields %}
- {{ field }}
{% endfor %}

To complete this task, follow these steps:

1. Analyze the news article and the extracted event carefully.
2. For each actor in the provided list, determine their role in the event based on the information in the news article. Note that some actors may not be involved in the event at all, in which case, simply ignore those.
3. Assign each actor to the most appropriate event argument.
4. If an event argument doesn't have a corresponding actor, leave it as an empty list.

Output the completed event arguments with the assigned actors in the given JSON format. Note that you should always include the full name of the actor.
```

Table 8: Prompt used for assigning entities to their correct event argument. A Pydantic schema is also given to the model to follow.

```
# instruction
You are an AI assistant tasked with extracting event arguments from a
    given news article. You will be provided with annotation guidelines
    for an event type and a news article to analyze.

# input
{{ article }}.
Extract the arguments of the main event in this article, which is of type
     {{ event_type }}.
For "entity" arguments, note that an entity can be a generic term like "
    Students" or "Protestors", or specific political groups, militia,
    armed groups, etc. Never use an individual's name as an entity.

Sometimes, a specific entity is accompanied with a generic entity. For
    example if a political party is leading a protest, both the political
     party's name and the "Protestors" should be included as entities.
When identifying an entity, provide as much information as possible.
```

Table 9: Prompt used for EAE.

C Hyperparameters

All fine-tuned models were fine-tuned with batch size 128 for 5 epochs. The final model checkpoint was selected for evaluation. Learning rate of 2×10^{-5}, cosine learning rate scheduler and AdamW (Loshchilov & Hutter, 2017) optimizer are used.

Training is done on a machine with 4 NVIDIA A100 GPU with 80GB, using DeepSpeed (Rasley et al., 2020) and the Transformers (Wolf et al., 2019) library.

For access to GPT-4o model, we used the OpenAI API. For access to OpenStreetMap, we used the publicly hosted version via Nominatim `https://nominatim.openstreetmap.org/`

D Full Schema of Lemonade

The following is the full schema of Lemonade, after conversion to Python code, in Pydantic (Colvin et al., 2024) format. Abstract classes (denoted by `ABC` are only meant to group event types together and store common event arguments, are not counted as an event type, and are not used by Zest. Docstrings are modified from the ACLED codebook (ACLED, 2023). `WomenTargetedCategory` and `Location` are two event types.

```
class Battle(ACLEDEvent, ABC):
    """
    A "Battle" event is defined as a violent interaction between two organized armed groups at a particular
      time and location. "Battle" can occur between armed and organized state, non-state, and external groups,
       and in any combination therein. There is no fatality minimum necessary for inclusion. Civilians can be
      harmed in the course of larger "Battle" events if they are caught in the crossfire, for example, or
      affected by strikes on military targets, which is commonly referred to as "collateral damage" (for more,
       see Indirect Killing of Civilians). When civilians are harmed in a "Battle" event, they are not
      recorded as an "Actor", nor is a separate civilian-specific event recorded. If any civilian fatalities
      are reported as part of a battle, they are aggregated in the "Fatalities" field for the "Battle" event.
    The specific elements of the definition of a "Battle" event are as follows:
    Violent interaction: the exchange of armed force, or the use of armed force at close distance, between
      armed groups capable of inflicting harm upon the opposing side.
    Organized armed groups: collective actors assumed to be operating cohesively around an agenda, identity,
      or political purpose, using weapons to inflict harm. These groups frequently have a designated name and
      stated agenda.
    The "Battle" event type may include: ground clashes between different armed groups, ground clashes between
       armed groups supported by artillery fire or airstrikes, ambushes of on-duty soldiers or armed militants
      , exchanges of artillery fire, ground attacks against military or militant positions, air attacks where
      ground forces are able to effectively fire on the aircraft, and air-to-air combat.
    Cases where territory is regained or overtaken without resistance or armed interaction are not recorded as
       "Battle" events. Instead, they are recorded as "NonStateActorOvertakesTerritory" under the "
      StrategicDevelopment" event type
    "Battle" event type has the following subtypes:
    - GovernmentRegainsTerritory: Government forces or their affiliates regain control of a location from
      competing state forces or non-state groups through armed interaction.
    - NonStateActorOvertakesTerritory: A non-state actor or foreign state actor captures territory from an
      opposing government or non-state actor through armed interaction, establishing a monopoly of force
      within that territory.
```

```
    - ArmedClash: Armed, organized groups engage in a battle without significant changes in territorial
      control.

    """

    location: Location = Field(..., description="Location where the event takes place")
    fatalities: Optional[int] = Field(
        ...,
        description="Total number of fatalities, if known",
    )


class GovernmentRegainsTerritory(Battle):
    """
    Is a type of "Battle" event. This event type is used when government forces or their affiliates that are
      fighting against competing state forces or against a non-state group regain control of a location
      through armed interaction. This event type is only recorded for the re-establishment of government
      control and not for cases where competing non-state actors exchange control. Short-lived and/or small-
      scale territorial exchanges that do not last for more than one day are recorded as "ArmedClash".
    """

    government_force: List[str] = Field(
        ...,
        description="The government forces or their affiliates that regain control of the territory",
    )
    adversary: List[str] = Field(
        ...,
        description="The competing state forces or non-state group that lose control of the territory. Can be
      State Forces, Rebel Groups, Political Militias, Identity Militias or External Forces",
    )


class NonStateActorOvertakesTerritory(Battle):
    """
    Is a type of "Battle" event. This event type is used when a non-state actor (excluding those operating
      directly on behalf of the government) or a foreign state actor, through armed interaction, captures
      territory from an opposing government or non-state actor; as a result, they are regarded as having a
      monopoly of force within that territory. Short-lived and/or small-scale territorial exchanges that do
      not last for more than one day are recorded as "ArmedClash" events. In cases where non-state forces
      fight with opposing actors in a location many times before gaining control, only the final territorial
      acquisition is recorded as "Non-state actor overtakes territory". All other battles in that location are
       recorded as "ArmedClash".
    """

    non_state_actor: List[str] = Field(
        ...,
        description="The non-state actor overtaking territory. Can be Rebel Groups, Political Militias,
      Identity Militias or External Forces",
    )
    adversary: List[str] = Field(
        ...,
        description="The opposing government or non-state actor from whom the territory was taken. Can be
      State Forces, Rebel Groups, Political Militias, Identity Militias or External Forces",
    )



class ArmedClash(Battle):
    """
    Is a type of "Battle" event. This event type is used when two organized groups like State Forces, Rebel
      Groups, Political Militias, Identity Militias or External Forces engage in a battle, and no reports
      indicate a significant change in territorial control.
    `side_1` and `side_2` denote the two sides of the armed clash.
    Excludes demonstrations that turn violent, riots, and other forms of violence that are not organized armed
       clashes.
    """


    side_1: List[str] = Field(
        ...,
        description="Groups involved in the clash. Can be State Forces, Rebel Groups, Political Militias,
      Identity Militias or External Forces",
    )
    side_2: List[str] = Field(
        ...,
        description="Groups involved in the clash. Can be State Forces, Rebel Groups, Political Militias,
      Identity Militias or External Forces",
    )
    targets_local_administrators: bool = Field(
        ...,
        description="Whether this violence is affecting local government officials and administrators -
      including governors, mayors, councilors, and other civil servants.",
    )
    women_targeted: List[WomenTargetedCategory] = Field(
        ...,
        description="The category of violence against women, if any. If this violence is not targeting women,
      this should be an empty list.",
    )



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. Events include
      individuals and groups who peacefully demonstrate against a political entity, government institution,
      policy, group, tradition, business, or other private institution. The following are not recorded as "
      Protest" events: symbolic public acts such as displays of flags or public prayers (unless they are
      accompanied by a demonstration); legislative protests, such as parliamentary walkouts or members of
      parliaments staying silent; strikes (unless they are accompanied by a demonstration); and individual
      acts such as self-harm actions like individual immolations or hunger strikes.
    Protestor are noted by generic actor name "Protestor". If they are representing a group, the name of that
      group is also recorded in the field.
    "Protest" event type has the following subtypes:
    - ExcessiveForceAgainstProtestors: Peaceful protestor are targeted with lethal violence or violence
      resulting in serious injuries by state or non-state actors.
    - ProtestWithIntervention: A peaceful protest is physically dispersed or suppressed without serious
      injuries, or protestor interact with armed groups or rioters without serious harm, or protestors are
      arrested.
    - PeacefulProtest: Demonstrators gather for a protest without engaging in violence or rioting and are not
      met with force or intervention.

    """

    location: Location = Field(..., description="Location where the event takes place")
    protestors: List[str] = Field(
        ...,
        description="List of protestor groups or individuals involved in the protest",
    )


class ExcessiveForceAgainstProtestors(Protest):
    """
    Is a type of "Protest" event (Protest events include individuals and groups who peacefully demonstrate
      against a political entity, government institution, policy, group, tradition, business, or other private
       institution.) This event type is used when individuals are engaged in a peaceful protest and are
      targeted with lethal violence or violence resulting in serious injuries (e.g. requiring hospitalization)
      . This includes situations where remote explosives, such as improvised explosive devices, are used to
      target protestors, as well as situations where non-state actors, such as rebel groups, target protestors
      .
    """

    # Possible "Interaction" codes include: 16, 26, 36, 46, 56, and 68.

    perpetrators: List[str] = Field(
        ...,
        description="Entities perpetrating the violence. Can be State Forces, Rebel Groups, Political Militias
      , Identity Militias, External Forces",
    )
    targets_civilians: bool = Field(
        ...,
        description="Indicates if the 'ExcessiveForceAgainstProtestors' event is mainly or only targeting
      civilians. E.g. state forces using lethal force to disperse peaceful protestors.",
    )

    fatalities: Optional[int] = Field(
        ...,
        description="Total number of fatalities, if known",
    )


class ProtestWithIntervention(Protest):
    """
    Is a type of "Protest" event. This event type is used when individuals are engaged in a peaceful protest
      during which there is a physically violent attempt to disperse or suppress the protest, which resulted
      in arrests, or minor injuries . If there is intervention, but not violent, the event is recorded as "
      PeacefulProtest" event type.
    """

    perpetrators: List[str] = Field(
        ...,
        description="Group(s) or entities attempting to disperse or suppress the protest",
    )
    fatalities: Optional[int] = Field(
        ...,
        description="Total number of fatalities, if known",
    )


class PeacefulProtest(Protest):
    """
    Is a type of "Protest" event (Protest events include individuals and groups who peacefully demonstrate
      against a political entity, government institution, policy, group, tradition, business, or other private
       institution.) This event type is 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.
    """

    # Possible "Interaction" codes include: 60, 66, and 67.

    counter_protestors: List[str] = Field(
```

```
        ..., description="Groups or entities engaged in counter protest, if any"
    )


class Riot(ACLEDEvent, ABC):
    """
    "Riot" are violent events where demonstrators or mobs of three or more engage in violent or destructive
      acts, including but not limited to physical fights, rock throwing, property destruction, etc. They may
      engage individuals, property, businesses, other rioting groups, or armed actors. Rioters are noted by
      generic actor name "Rioters". If rioters are affiliated with a specific group - which may or may not be
      armed - or identity group, that group is recorded in the respective "Actor" field. Riots may begin as
      peaceful protests, or a mob may have the intention to engage in violence from the outset.
    "Riot" event type has the following subtypes:
    - ViolentDemonstration: Demonstrators engage in violence or destructive activities, such as physical
      clashes, vandalism, or road-blocking, regardless of who initiated the violence.
    - MobViolence: Rioters violently interact with other rioters, civilians, property, or armed groups outside
       of demonstration contexts, often involving disorderly crowds with the intention to cause harm or
      disruption.

    """

    location: Location = Field(..., description="Location where the event takes place")
    fatalities: Optional[int] = Field(
        ...,
        description="Total number of fatalities, if known",
    )
    targets_civilians: bool = Field(
        ...,
        description="Indicates if the 'Riot' event is mainly or only targeting civilians. E.g. a village mob
      assaulting another villager over a land dispute.",
    )
    group_1: List[str] = Field(
        ..., description="Group or individual involved in the violence"
    )
    group_2: List[str] = Field(
        ...,
        description="The other group or individual involved in the violence, if any",
    )
    targets_local_administrators: bool = Field(
        ...,
        description="Whether this violence is affecting local government officials and administrators -
      including governors, mayors, councilors, and other civil servants.",
    )
    women_targeted: List[WomenTargetedCategory] = Field(
        ...,
        description="The category of violence against women, if any. If this violence is not targeting women,
      this should be an empty list.",
    )


class ViolentDemonstration(Riot):
    """
    Is a type of "Riot" event. This event type is used when demonstrators engage in violence and/or
      destructive activity. Examples include physical clashes with other demonstrators or government forces;
      vandalism; and road-blocking using barricades, burning tires, or other material. The coding of an event
      as a "Violent demonstration" does not necessarily indicate that demonstrators initiated the violence and
      /or destructive actions.
    Excludes events where a weapon is drawn but not used, or when the situation is de-escalated before
      violence occurs.
    """


class MobViolence(Riot):
    """
    Is a type of "Riot" event. A mob is considered a crowd of people that is disorderly and has the intention
      to cause harm or disruption through violence or property destruction. Note that this type of violence
      can also include spontaneous vigilante mobs clashing with other armed groups or attacking civilians.
      While a "Mob violence" event often involves unarmed or crudely armed rioters, on rare occasions, it can
      involve violence by people associated with organized groups and/or using more sophisticated weapons,
      such as firearms.
    """


class ExplosionOrRemoteViolence(ACLEDEvent, ABC):
    """
    "ExplosionOrRemoteViolence" is defined as events as incidents in which one side uses weapon types that, by
       their nature, are at range and widely destructive. The weapons used in "ExplosionOrRemoteViolence"
      events are explosive devices, including but not limited to: bombs, grenades, improvised explosive
      devices (IEDs), artillery fire or shelling, missile attacks, air or drone strikes, and other widely
      destructive heavy weapons or chemical weapons. Suicide attacks using explosives also fall under this
      category. When an "ExplosionOrRemoteViolence" event is reported in the context of an ongoing battle, it
      is merged and recorded as a single "Battles" event. "ExplosionOrRemoteViolence" can be used against
      armed agents as well as civilians.
    "ExplosionOrRemoteViolence" event type has the following subtypes:
    - ChemicalWeapon: The use of chemical weapons in warfare without any other engagement.
    - AirOrDroneStrike: Air or drone strikes occurring without any other engagement, including attacks by
      helicopters.
```

```
    - SuicideBomb: A suicide bombing or suicide vehicle-borne improvised explosive device (SVBIED) attack
      without an armed clash.
    - ShellingOrArtilleryOrMissileAttack: The use of long-range artillery, missile systems, or other heavy
      weapons platforms without any other engagement.
    - RemoteExplosiveOrLandmineOrIED: Detonation of remotely- or victim-activated devices, including landmines
       and IEDs, without any other engagement.
    - Grenade: The use of a grenade or similar hand-thrown explosive without any other engagement.
    """

    location: Location = Field(..., description="Location where the event takes place")
    targets_civilians: bool = Field(
        ...,
        description="Indicates if the 'ExplosionOrRemoteViolence' event is mainly or only targeting civilians.
     E.g. a landmine killing a farmer.",
    )
    fatalities: Optional[int] = Field(
        ...,
        description="Total number of fatalities, if known",
    )
    attackers: List[str] = Field(..., description="Entities conducting the violence")
    targeted_entities: List[str] = Field(
        ..., description="Entities or actors being targeted"
    )
    targets_local_administrators: bool = Field(
        ...,
        description="Whether this violence is affecting local government officials and administrators -
      including governors, mayors, councilors, and other civil servants.",
    )
    women_targeted: List[WomenTargetedCategory] = Field(
        ...,
        description="The category of violence against women, if any. If this violence is not targeting women,
      this should be an empty list.",
    )


class ChemicalWeapon(ExplosionOrRemoteViolence):
    """
    Is a type of "ExplosionOrRemoteViolence" event. This event type captures the use of chemical weapons in
      warfare in the absence of any other engagement. ACLED considers chemical weapons as all substances
      listed as Schedule 1 of the Chemical Weapons Convention, including sarin gas, mustard gas, chlorine gas,
       and anthrax. Napalm and white phosphorus, as well as less-lethal crowd control substances - such as
      tear gas - are not considered chemical weapons within this event type.
    """


class AirOrDroneStrike(ExplosionOrRemoteViolence):
    """
    Is a type of "ExplosionOrRemoteViolence" event. This event type is used when air or drone strikes take
      place in the absence of any other engagement. Please note that any air-to-ground attacks fall under this
       event type, including attacks by helicopters that do not involve exchanges of fire with forces on the
      ground.
    """


class SuicideBomb(ExplosionOrRemoteViolence):
    """
    Is a type of "ExplosionOrRemoteViolence" event. This event type is used when a suicide bombing occurs in
      the absence of an armed clash, such as an exchange of small arms fire with other armed groups. It also
      includes suicide vehicle-borne improvised explosive device (SVBIED) attacks. Note that the suicide
      bomber is included in the total number of reported fatalities coded for such events.
    """


class ShellingOrArtilleryOrMissileAttack(ExplosionOrRemoteViolence):
    """
    Is a type of "ExplosionOrRemoteViolence" event. This event type captures the use of long-range artillery,
      missile systems, or other heavy weapons platforms in the absence of any other engagement. When two armed
       groups exchange long-range fire, it is recorded as an "ArmedClash". "ShellingOrArtilleryOrMissileAttack
     " events include attacks described as shelling, the use of artillery and cannons, mortars, guided
      missiles, rockets, grenade launchers, and other heavy weapons platforms. Crewed aircraft shot down by
      long-range systems fall under this event type.  Uncrewed armed drones that are shot down, however, are
      recorded as interceptions under "DisruptedWeaponsUse" because people are not targeted (see below).
      Similarly, an interception of a missile strike itself (such as by the Iron Dome in Israel) is also
      recorded as "DisruptedWeaponsUse".
    """


class RemoteExplosiveOrLandmineOrIED(ExplosionOrRemoteViolence):
    """
    Is a type of "ExplosionOrRemoteViolence" event. This event type is used when remotely- or victim-activated
       devices are detonated in the absence of any other engagement. Examples include landmines, IEDs -
      whether alone or attached to a vehicle, or any other sort of remotely detonated or triggered explosive.
      Unexploded ordnances (UXO) also fall under this category.
    SVBIEDs are recorded as "Suicide bomb" events, while the safe defusal of an explosive or its accidental
      detonation by the actor who planted it (with no other casualties reported) is recorded under "
      DisruptedWeaponsUse".
    """
```

```
class Grenade(ExplosionOrRemoteViolence):
    """
    Is a type of "ExplosionOrRemoteViolence" event. This event type captures the use of a grenade or any other
       similarly hand-thrown explosive, such as an IED that is thrown, in the absence of any other engagement.
       Events involving so-called "crude bombs" (such as Molotov cocktails, firecrackers, cherry bombs, petrol
       bombs, etc.) as well as "stun grenades" are not recorded in this category, but are included under
      either "Riot" or "StrategicDevelopment" depending on the context in which they occurred.
    """


class ViolenceAgainstCivilians(ACLEDEvent, ABC):
    """
    ACLED defines "ViolenceAgainstCivilians" as violent events where an organized armed group inflicts
      violence upon unarmed non-combatants. By definition, civilians are unarmed and cannot engage in
      political violence. Therefore, the violence is understood to be asymmetric as the perpetrator is assumed
       to be the only actor capable of using violence in the event. The perpetrators of such acts include
      state forces and their affiliates, rebels, militias, and external/other forces.
    In cases where the identity and actions of the targets are in question (e.g. the target may be employed as
       a police officer), ACLED determines that if a person is harmed or killed while unarmed and unable to
      either act defensively or counter-attack, this is an act of "ViolenceAgainstCivilians". This includes
      extrajudicial killings of detained combatants or unarmed prisoners of war.
    "ViolenceAgainstCivilians" also includes attempts at inflicting harm (e.g. beating, shooting, torture,
      rape, mutilation, etc.) or forcibly disappearing (e.g. kidnapping and disappearances) civilian actors.
      Note that the "ViolenceAgainstCivilians" event type exclusively captures violence targeting civilians
      that does not occur concurrently with other forms of violence - such as rioting - that are coded higher
      in the ACLED event type hierarchy. To get a full list of events in the ACLED dataset where civilians
      were the main or only target of violence, users can filter on the "Civilian targeting" field.
    "ViolenceAgainstCivilians" event type has the following subtypes:
    - SexualViolence: Any event where an individual is targeted with sexual violence, including but not
      limited to rape, public stripping, and sexual torture, with the gender identities of victims recorded
      when reported.
    - Attack: An event where civilians are targeted with violence by an organized armed actor outside the
      context of other forms of violence, including severe government overreach by law enforcement.
    - AbductionOrForcedDisappearance: An event involving the abduction or forced disappearance of civilians
      without reports of further violence, including arrests by non-state groups and extrajudicial detentions
      by state forces, but excluding standard judicial arrests by state forces.
    """

    location: Location = Field(..., description="Location where the event takes place")
    targets_local_administrators: bool = Field(
        ...,
        description="Whether this violence is affecting local government officials and administrators -
      including governors, mayors, councilors, and other civil servants.",
    )
    women_targeted: List[WomenTargetedCategory] = Field(
        ...,
        description="The category of violence against women, if any. If this violence is not targeting women,
      this should be an empty list.",
    )


class SexualViolence(ViolenceAgainstCivilians):
    """
    Is a type of "ViolenceAgainstCivilians" event. This event type is used when any individual is targeted
      with sexual violence. SexualViolence is defined largely as an action that inflicts harm of a sexual
      nature. This means that it is not limited to solely penetrative rape, but also includes actions like
      public stripping, sexual torture, etc. Given the gendered nature of sexual violence, the gender
      identities of the victims - i.e. "Women", "Men", and "LGBTQ\+", or a combination thereof - are recorded
      in the "Associated Actor" field for these events when reported. Note that it is possible for sexual
      violence to occur within other event types such as "Battle" and "Riot".
    """

    fatalities: Optional[int] = Field(
        ...,
        description="Total number of fatalities, if known",
    )  # Is very very rare, only 7 events in English for 2024
    perpetrators: List[str] = Field(..., description="The attacker(s) entity or actor")
    victims: List[str] = Field(
        ...,
        description="The entity or actor(s) that is the target or victim of the SexualViolence event",
    )


class Attack(ViolenceAgainstCivilians):
    """
    Is a type of "ViolenceAgainstCivilians" event. This event type is used when civilians are targeted with
      violence by an organized armed actor outside the context of other forms of violence like ArmedClash,
      Protests, Riots, or ExplosionOrRemoteViolence. Violence by law enforcement that constitutes severe
      government overreach is also recorded as an "Attack" event.
    Attacks of a sexual nature are recorded as SexualViolence.
    If only property is attacked and not people, the event should be recorded as LootingOrPropertyDestruction
      event type.
    Excludes discovery of mass graves, which are recorded as "OtherStrategicDevelopment" events.
    """

    fatalities: Optional[int] = Field(
```

```
        ...,
        description="Total number of fatalities, if known",
    )
    attackers: List[str] = Field(..., description="The attacker entity or actor(s)")
    targeted_entities: List[str] = Field(
        ..., description="The entity or actor(s) that is the target of the attack"
    )


class AbductionOrForcedDisappearance(ViolenceAgainstCivilians):
    """
    Is a type of "ViolenceAgainstCivilians" event. This event type is used when an actor engages in the
      abduction or forced disappearance of civilians, without reports of further violence. If fatalities or
      serious injuries are reported during the abduction or forced disappearance, the event is recorded as an
      "Attack" event instead. If such violence is reported in later periods during captivity, this is recorded
       as an additional "Attack" event. Note that multiple people can be abducted in a single "Abduction/
      forced disappearance" event.
    Arrests by non-state groups and extrajudicial detentions by state forces are considered "Abduction/forced
      disappearance". Arrests conducted by state forces within the standard judicial process are, however,
      considered "Arrest".
    """

    abductor: List[str] = Field(..., description="The abductor person or group(s)")
    abductee: List[str] = Field(
        ...,
        description="People or group(s) that were abducted or disappeared. Note that multiple people can be
      abducted in a single AbductionOrForcedDisappearance event",
    )


class StrategicDevelopment(ACLEDEvent, ABC):
    """
    This event type captures contextually important information regarding incidents and activities of groups
      that are not recorded as "Political violence" or "Demonstration" events, yet may trigger future events
      or contribute to political dynamics within and across states. The inclusion of such events is limited,
      as their purpose is to capture pivotal events within the broader political landscape. They typically
      include a disparate range of events, such as recruitment drives, looting, and incursions, as well as the
       location and date of peace talks and the arrests of high-ranking officials or large groups. While it is
       rare for fatalities to be reported as a result of such events, they can occur in certain cases - e.g.
      the suspicious death of a high-ranking official, the accidental detonation of a bomb resulting in the
      bomber being killed, etc.
    Due to their context-specific nature, "StrategicDevelopment" are not collected and recorded in the same
      cross-comparable fashion as "Political violence" and "Demonstration" events. As such, the "
      StrategicDevelopment" event type is primarily a tool for understanding particular contexts.
    "StrategicDevelopment" event type has the following subtypes:
    - Agreement: Records any agreement between different actors, such as peace talks, ceasefires, or prisoner
      exchanges.
    - Arrest: Used when state forces or controlling actors detain a significant individual or conduct
      politically important mass arrests.
    - ChangeToArmedGroup: Records significant changes in the activity or structure of armed groups, including
      creation, recruitment, movement, or absorption of forces.
    - DisruptedWeaponsUse: Captures instances where an explosion or remote violence event is prevented, or
      when significant weapons caches are seized.
    - BaseEstablished: Used when an organized armed group establishes a permanent or semi-permanent base or
      headquarters.
    - LootingOrPropertyDestruction: Records incidents of looting or seizing goods/property outside the context
       of other forms of violence or destruction.
    - NonViolentTransferOfTerritory: Used when actors acquire control of a location without engaging in
      violent interaction with another group.
    - OtherStrategicDevelopment: Covers significant developments that don't fall into other Strategic
      Development event types, such as coups or population displacements.
    """

    location: Location = Field(..., description="Location where the event takes place")


class Agreement(StrategicDevelopment):
    """
    Is a type of "StrategicDevelopment" event. This event type is used to record any sort of agreement between
       different armed actors (such as governments and rebel groups). Examples include peace agreements/talks,
       ceasefires, evacuation deals, prisoner exchanges, negotiated territorial transfers, prisoner releases,
      surrenders, repatriations, etc.
    Excludes agreements between political parties, trade unions, or other non-armed actors like protestors.
    """

    group_1: List[str] = Field(
        ..., description="Group or individual involved in the agreement"
    )
    group_2: List[str] = Field(
        ...,
        description="The other group or individual involved in the agreement",
    )


class Arrest(StrategicDevelopment):
    """
    Is a type of "StrategicDevelopment" event. This event type is used when state forces or other actors
      exercising de facto control over a territory either detain a particularly significant individual or
```

```
      engage in politically significant mass arrests. This excludes arrests of individuals for common crimes,
      such as theft or assault, unless the individual is a high-ranking official or the arrest is politically
      significant.
    """

    detainers: List[str] = Field(
        ..., description="The person or group(s) who detains or jails the detainee(s)"
    )
    detainees: List[str] = Field(
        ..., description="The person or group(s) being detained or jailed"
    )


class ChangeToArmedGroup(StrategicDevelopment):
    """
    Is a type of "StrategicDevelopment" event. This event type is used to record significant changes in the
      activity or structure of armed groups. It can cover anything from the creation of a new rebel group or a
       paramilitary wing of the security forces, "voluntary" recruitment drives, movement of forces, or any
      other non-violent security measures enacted by armed actors. This event type can also be used if one
      armed group is absorbed into a different armed group or to track large-scale defections.
    """

    armed_group: List[str] = Field(
        ..., description="The name of armed group that underwent change"
    )
    other_actors: List[str] = Field(
        ...,
        description="Other actors or groups involved. E.g. the government that ordered a change to its army.",
    )


class DisruptedWeaponsUse(StrategicDevelopment):
    """
    Is a type of "StrategicDevelopment" event. This event type is used to capture all instances in which an
      event of "ExplosionOrRemoteViolence" is prevented from occurring, or when armed actors seize significant
       caches of weapons. It includes the safe defusal of an explosive, the accidental detonation of
      explosives by those allegedly responsible for planting it, the interception of explosives in the air, as
       well as the seizure of weapons or weapons platforms such as jets, helicopters, tanks, etc. Note that in
       cases where a group other than the one that planted an explosive is attempting to render an explosive
      harmless and it goes off, this is recorded under the "ExplosionOrRemoteViolence" event type, as the
      explosive has harmed an actor other than the one that planted it.
    """

    attackers: List[str] = Field(
        ..., description="The entity or actor(s) responsible for the remote violence"
    )
    disruptors: List[str] = Field(
        ...,
        description="The entity or actor(s) disrupting the explosion or remote violence",
    )
    targets_local_administrators: bool = Field(
        ...,
        description="Whether this violence is affecting local government officials and administrators -
      including governors, mayors, councilors, and other civil servants.",
    )
    women_targeted: List[WomenTargetedCategory] = Field(
        ...,
        description="The category of violence against women, if any. If this violence is not targeting women,
      this should be an empty list.",
    )


class BaseEstablished(StrategicDevelopment):
    """
    Is a type of "StrategicDevelopment" event. This event type is used when an organized armed group
      establishes a permanent or semi-permanent base or headquarters. There are few cases where opposition
      groups other than rebels can also establish a headquarters or base (e.g. AMISOM forces in Somalia).
    """

    group: List[str] = Field(
        ..., description="Entity or group(s) establishing the base"
    )


class LootingOrPropertyDestruction(StrategicDevelopment):
    """
    Is a type of "StrategicDevelopment" event. This event type is used when actors engage in looting or
      seizing goods or property outside the context of other forms of violence or destruction, such as rioting
       or armed clashes. This excludes the seizure or destruction of weapons or weapons systems, which are
      captured under the "DisruptedWeaponsUse" event type. This can occur during raiding or after the capture
      of villages or other populated places by armed groups that occur without reported violence.
    """

    perpetrators: List[str] = Field(
        ..., description="The group or entity that does the looting or seizure"
    )
    victims: List[str] = Field(
        ..., description="The group or entity that was the target of looting or seizure"
```

```python
    )
    targets_local_administrators: bool = Field(
        ...,
        description="Whether this violence is affecting local government officials and administrators -
      including governors, mayors, councilors, and other civil servants.",
    )
    women_targeted: List[WomenTargetedCategory] = Field(
        ...,
        description="The category of violence against women, if any. If this violence is not targeting women,
      this should be an empty list.",
    )


class NonViolentTransferOfTerritory(StrategicDevelopment):
    """
    Is a type of "StrategicDevelopment" event. This event type is used in situations in which rebels,
      governments, or their affiliates acquire control of a location without engaging in a violent interaction
       with another group. Rebels establishing control of a location without any resistance is an example of
      this event.
    """

    actors_taking_over: List[str] = Field(
        ..., description="The entity or actor(s) establishing control."
    )
    actors_giving_up: List[str] = Field(
        ..., description="The entity or actor(s) giving up territory, if known."
    )


class OtherStrategicDevelopment(StrategicDevelopment):
    """
    Is a type of "StrategicDevelopment" event. This event type is used to cover any significant development
      that does not fall into any of the other "StrategicDevelopment" event types. Includes the occurrence of
      a coup, the displacement of a civilian population as a result of fighting, and the discovery of mass
      graves.
    """

    group_1: List[str] = Field(
        ..., description="Group or individual involved in the StrategicDevelopment"
    )
    group_2: List[str] = Field(
        ...,
        description="The other group or individual involved in the violence, if any",
    )


class WomenTargetedCategory(str, Enum):
    CANDIDATES_FOR_OFFICE = "Women who are running in an election to hold a publicly elected government
      position"
    POLITICIANS = "Women who currently serve in an elected position in government"
    POLITICAL_PARTY_SUPPORTERS = "political party supporters"
    VOTERS = "Women who are registering to vote or are casting a ballot in an election"
    GOVERNMENT_OFFICIALS = "Women who work for the local, regional, or national government in a non-partisan
      capacity"
    ACTIVISTS_HRD_SOCIAL_LEADERS = (
        "Women who are activists/human rights defenders/social leaders"
    )
    RELATIVES_OF_TARGETED_GROUPS = "Women who are subject to violence as a result of who they are married to,
      the daughter of, related to, or are otherwise personally connected to (e.g. candidates, politicians,
      social leaders, armed actors, voters, party supporters, etc.)"
    ACCUSED_OF_WITCHCRAFT = "Women accused of witchcraft or sorcery, or other mystical or spiritual practices
      that are typically considered taboo or dangerous within some societies (excluding women who serve as
      religious leaders in religious structures that are typically not viewed as taboo or dangerous, such as
      nuns, female priests, or shamans)"
    GIRLS = "Girls who are under the age of 18; they may be specifically referred to by age or explicitly
      referred to as a child/girl"


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 upto the
      neighborhood level, including village/city, district, county, province, region, country, if available.
      Exclude street names, buildings, and other specific landmarks.",
    )
```

The languages included in LEMONADE are in Table 10.

Table 10: Mapping of language acronyms.

Acronym	Full Name
en	English
es	Spanish
ar	Arabic
fr	French
pt	Portuguese
ko	Korean
de	German
uk	Ukrainian
my	Malay
it	Italian
tr	Turkish
id	Indonesian
ru	Russian
fa	Persian (Farsi)
ne	Nepali
zh	Chinese