

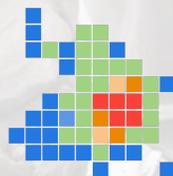
# From Scarcity to Signal

Combining LLMs, Synthetic Data and an Active Learning Framework for Rare Event Detection

Sonja Häffner<sup>1</sup> · Simon Polichinel von der Maase<sup>1</sup> · Xiaolong Sun<sup>2</sup> · Dylan Pinheiro<sup>1</sup> · Kristine Helskog<sup>1</sup>

<sup>1</sup> Peace Research Institute Oslo (PRIO) · <sup>2</sup> Uppsala University

**VIEWS**  
PREDICTING CONFLICT &  
HUMANITARIAN IMPACTS



UPPSALA  
UNIVERSITET



PRIO



## Kabul blast kills teenagers sitting practice exam

1 October 2022

Share Save



Watch: Relatives search for loved ones after Afghanistan blast

**A suicide attack at a tuition centre in the Afghan capital Kabul has killed at least 35 people, most of them female students, the United Nations says.**

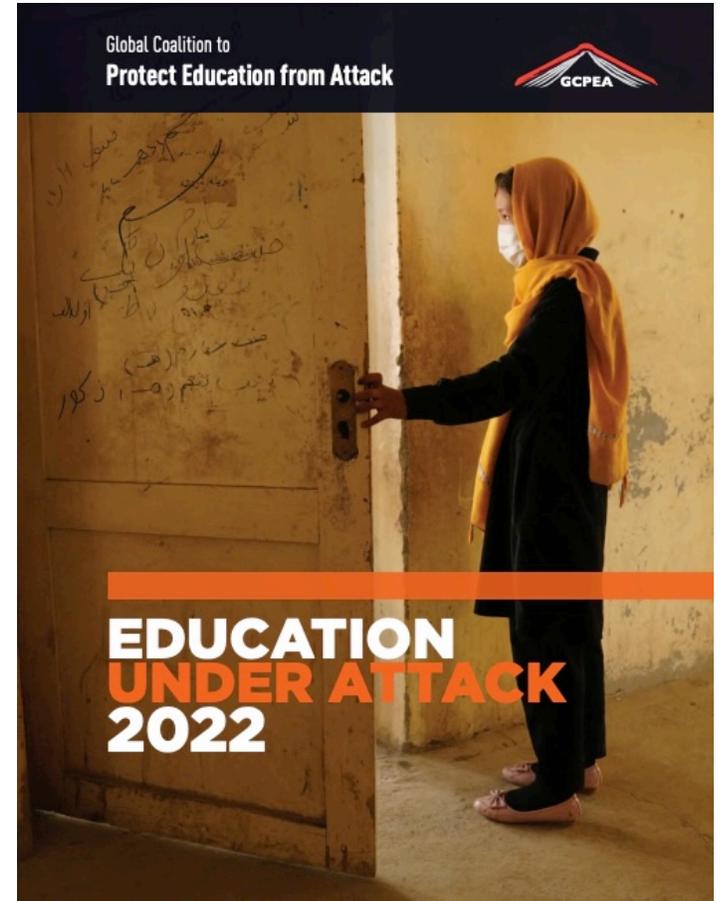
More than 80 others were wounded in Friday's attack at the Kaaj education centre in the Dasht-e-Barchi area.

## Attacks on Education (EdAttack)

*EdAttack* studies the impacts of attacks on children, short & long term

### Why it matters

- **+20%** increase in 2022–2023 (GCPEA)
- **6,000** attacks across 2022–2023
- **10,000+** students & educators harmed  $\approx$  **8 attacks/day**



**Yet, structured datasets (UCDP, ACLED) do not explicitly annotate this dimension of violence**

**This makes systematic study and monitoring impossible with existing data**

## **Dimensions of Attacks on Education:**

What we are interested in: The targeting of **educational settings** and **actors** *in their role as educators, educational personell or students.*

### **Core Modalities:**

<b>Violence</b>	<b>Threats &amp; Intimidation</b>	<b>Abductions &amp; Kidnapping</b>
<b>Military Occupation</b>	<b>Sexual Violence</b>	<b>Arrests &amp; Detention</b>

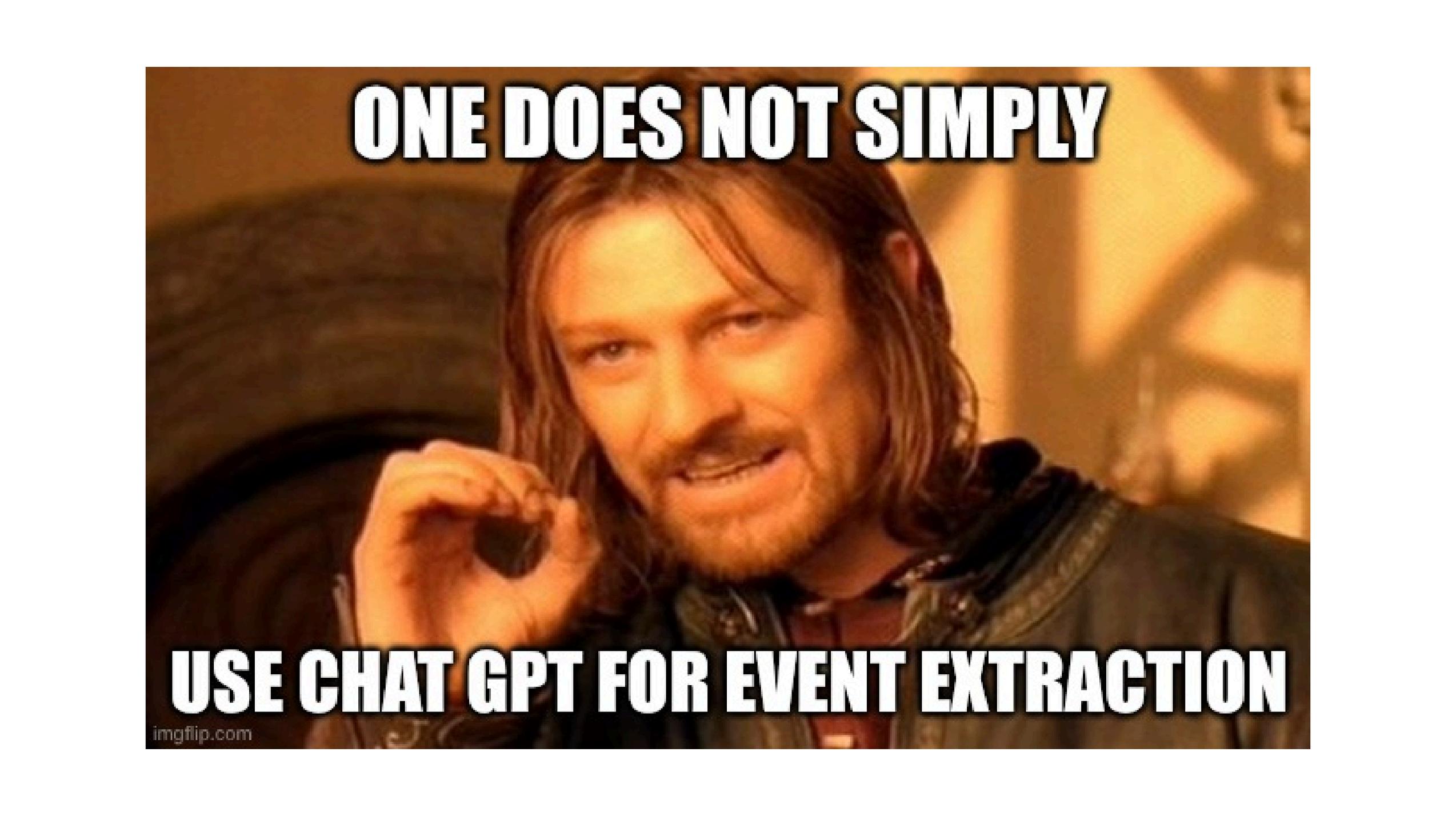
**Additional Flags:** level of education, gender dimensions, incidental, armed conflict actors mentioned, target



## Why Manual Annotation Isn't The Answer

- Each new relevant label (e.g. attacks on schools, students, personnel, abductions) **multiplies the operational** cost of manual annotation
- Manually re-annotating millions of past events and/or articles is **infeasible without vast amount of dedicated funding**
- One-off static datasets can be built, yes, but they quickly **become outdated and cannot support operational needs** such as monitoring or forecasting
- Humans will likely keep playing a pivotal role in event extraction, but we need a scalable solution that can help us **add more context and labels** to event data

**Why not just use Chat GPT, Claude, or Gemini?**

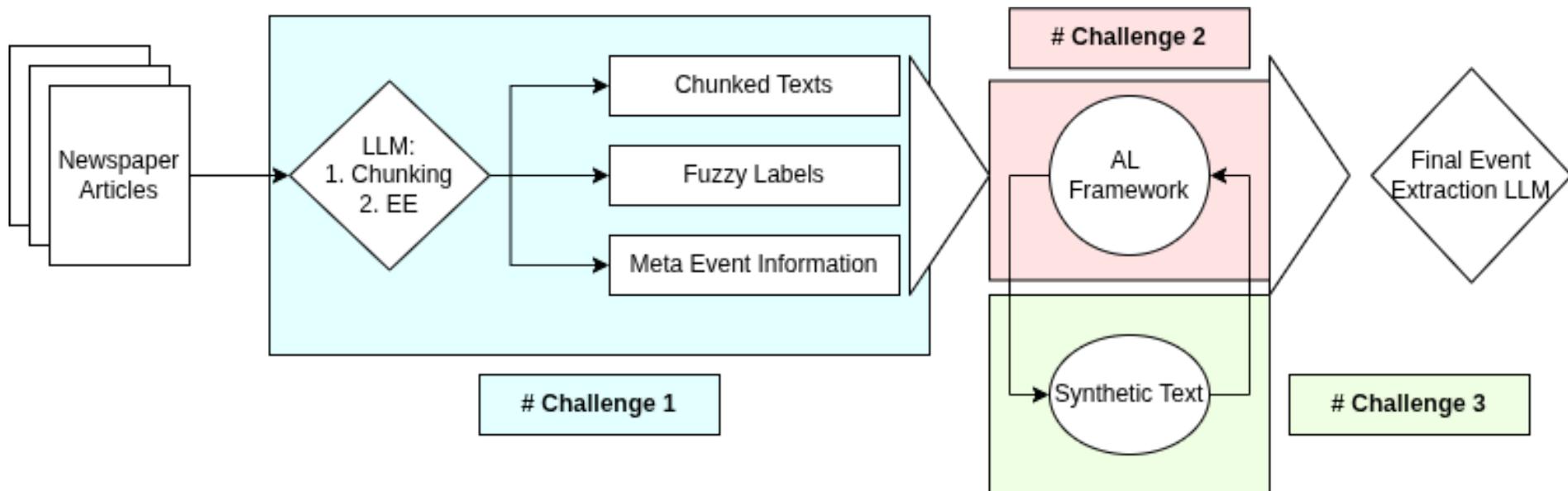


**ONE DOES NOT SIMPLY**

**USE CHAT GPT FOR EVENT EXTRACTION**

# ⚙️ How we address it: Building the Edattack Engine!

1. Relevant article selection from Factiva & preselection stage
2. Event extraction framework
3. Active learning framework: combining synthetic and real-life data
4. (Final event extraction layer)



## #Challenge 1: Event Extraction

- ACLED & UCDP are still considered the goldstandard and rely on human annotations (manual work)
- Humans struggle with event extraction! → low ICR amongst coders
- Our concept of interest (AoE) is nuanced, core modalities and flags are hierarchical  
→ more than a simple yes/no classification
- Instability of results depending on prompt and model chosen (especially for self hosted LLMs )
- **Struggles:** hallucination, multiple events, longer texts, prompt length
- **Solution:** Cascading prompts that 1. chop up an article in unique event spans, 2. use one prompt per core modality, 3. consolidates events and 4th assigns text pieces to it.
- **Human coders as quality assurance**

## Example Text

Universities have been a frequent flashpoint anti-government unrest in Iran and students say the murder shows authorities are not providing sufficient safety. "Tehran University dormitory is currently under siege ... More than 100 plainclothes officers are stationed at the entrance to Tehran University's dormitory," the Amirkabir university newsletter reported on its social media channels on Friday. The student newsletter reported that a student it named as Amir Hossein Gohari was beaten and detained by plainclothes forces, while three other students were arrested, but their names remain unknown. Earlier in the day, Tehran University students staged a protest against the university's policies and the lack of student safety in response to the death of their peer. Amir Mohammad Khaleghi, an undergraduate student at the university's business management department, was attacked by robbers earlier this week and died from his injuries. [...] The Amirkabir newsletter reported that some students plan to continue their protest in a sit in in the courtyard of the university until the morning despite the presence of security forces. [...] According to Iranian media, Khaleghi was attacked with a sharp weapon on Wednesday evening by thieves in a neighborhood near Tehran University's dormitories. [..]

## Extracted Event

Field	Value
Date	Friday
Location	Teheran University
Perpetrators	plainclothes officers
Targets	university
Summary	Plainclothes officers are currently stationed at Teheran University.
Core Modalities	Military occupation
Modality Triggers	under siege; stationed
Casualties	0
Sources	student newsletter
Span IDs	2, 11

## Problems

- **Definition of an event:** An AoE is a distinct real-world incident in which one identifiable actor performs one action that maps to exactly one core modality, occurring at a specific time and location (explicitly or implicitly). Multiple descriptions of the same incident are merged into one event, while actions belonging to different modalities are split into separate events.
- annotation unit?- only code present events (add a marker for it)?
- unit of analysis: What is an event? Self-hosted LLMs struggle
- Instability of coding depending on LLM & prompt
- Chat GPT is not good at this (also license problems)?
- ...

## Challenge 2: Active Learning Framework

- **Core idea:** To balance quality and scalability
- **Possible solution** query only the most informative examples → **fewer labels** for the same accuracy.
- **Classical result:** uncertainty sampling reduced labeled data needs by **up to 500×** on text classification. (Lewis & Gale, 1994)
- Even with **low levels of intercoder reliability**, AL performs better than random sampling (Miller et. al 2020)
- **Applied evidence:** up to **28–35%** less annotation time in clinical NLP. (Kholghi et al., 2017)

# Our Active Learning Framework

## AL-Loop

- Train → predict → select "most uncertain" → human annotators → repeat.

## Task & Model

- Model: ConflBERT
- Task: Multi-class, multi label classification

## Human-in-the-loop

- 4 coders
- Agreement metric: Krippendorff's alpha & Fleiss's Kappa (Jacard & MASI)
- Disagreement handling via majority vote and consensus enforcement in absence of majority

# **Our Active Learning Framework**

## **Querying strategies (per AL round)**

1. **Variance Learner** Highest variance
2. **Entropy Learner** Model's confidence on average
3. **BALD Learner** Model's internal uncertainty
4. **CoreSet Learner** Ensure text diversity
5. **Information Density Learner** Combines entropy and density (how common is this example)

## **Challenge 3: Enriching Rare Event Data with Synthetic Data**

Based on human-labelled data, we use language-model–based data augmentation to expand scarce labeled text.

Techniques:

- Synonym replacement → linguistic diversity (Wei and Zu 2019)
- Paraphrasing (Hedge and Patil 2020)
- Round-trip translation → cross-lingual robustness (Beddiar and Oussalah 2021, Kimera et al. 2024)
- Counterfactual generation → create alternative event outcomes (Madaan et al. 2020, Wu et al. 2021)

These augmentations expand limited high-quality data while keeping events realistic, improving model robustness and recall for rare signals.

## Examples

### Original:

3 girls and 2 soldiers killed in Pakistan school bus bombing are buried  
QUETTA, Pakistan (AP) — Hundreds of mourners in Pakistan on Thursday attended the funerals of three schoolgirls and two soldiers killed in a suicide bombing that targeted a school bus.

### Synonym:

3 females and 2 servicemen perished in Pakistan school bus blast are interred  
QUETTA, Pakistan (AP) — Hundreds of grievors in Pakistan on Thursday participated in the burials of three schoolgirls and two soldiers slain in a self-destructive detonation that struck a school bus.

### Paraphrase:

Funerals were held on Thursday in Pakistan for three schoolgirls and two soldiers who died in a school bus bombing, with hundreds in attendance.

### Backtranslation\_fr\_zh:

QUETTA, Pakistan (AP) — A large number of attendees gathered in Pakistan on Thursday for the burial services of three female students and two military personnel who died in a suicide attack on a school bus.

### Counterfactual:

QUETTA, Pakistan (AP) — Hundreds of mourners in Pakistan on Thursday attended the funerals of three students and two school staff members killed in a suicide bombing that targeted a school bus.

## Our Data

- 2.3 Mio. events from the **Armed Conflict Location and Event Dataset (ACLED)** (January 1997 to November 2024)
- Random train and test split 80/20
- After preselection: 28000 candidate events

### Example ACLED Notes and Corresponding Labels

ACLED note	Human labels
On 6 November 2024, Ambazonian Separatists from the Restoration Forces faction abducted a teacher from his private residence in Konye (Nguti, Kupe-Manenguba, Sud-Ouest) for encouraging parents to send their children to school.	<i>Recruitment &amp; Abduction; Target: Educational Personnel; Male-mentioned</i>
"Students destroyed university buildings after they were not allowed to hold a meeting on campus. Police used teargas against rioters and arrested 10 students."	<i>No Edattack</i>
"In the previous 2 weeks, 16 schools, including a communal secondary school, were set on fire by armed rebels in Rural Bujumbura."	<i>Violence; Target: Infrastructure; Educational-Level: Secondary</i>

## 💡 Our Active Learning Loop

### 🩺 Annotation & Reliability

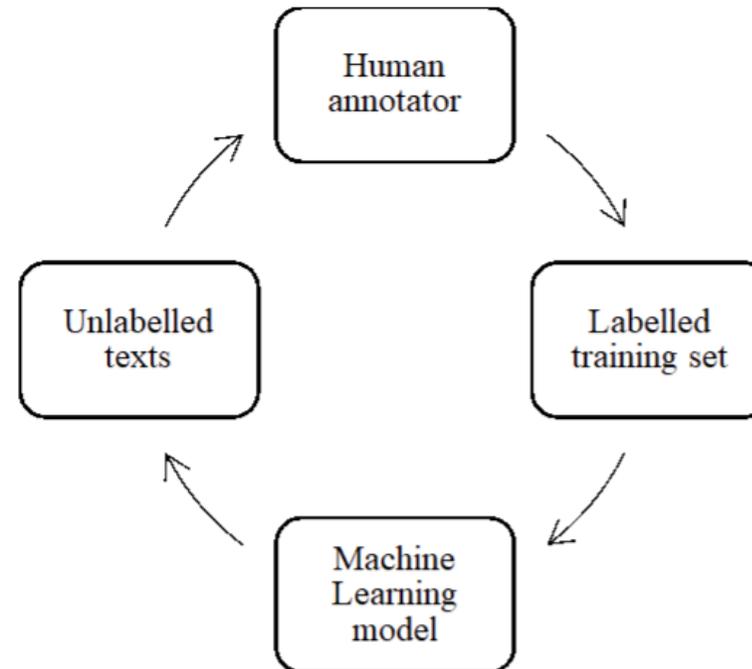
- 200 texts · 4 annotators
- Majority + consensus voting
- **Krippendorff's  $\alpha = 0.71$**

### 🔄 Iterative Learning

- 8 active learning rounds
- ~4,000 labeled texts (Doccano)

### 📊 Evaluation

- Holdout test set



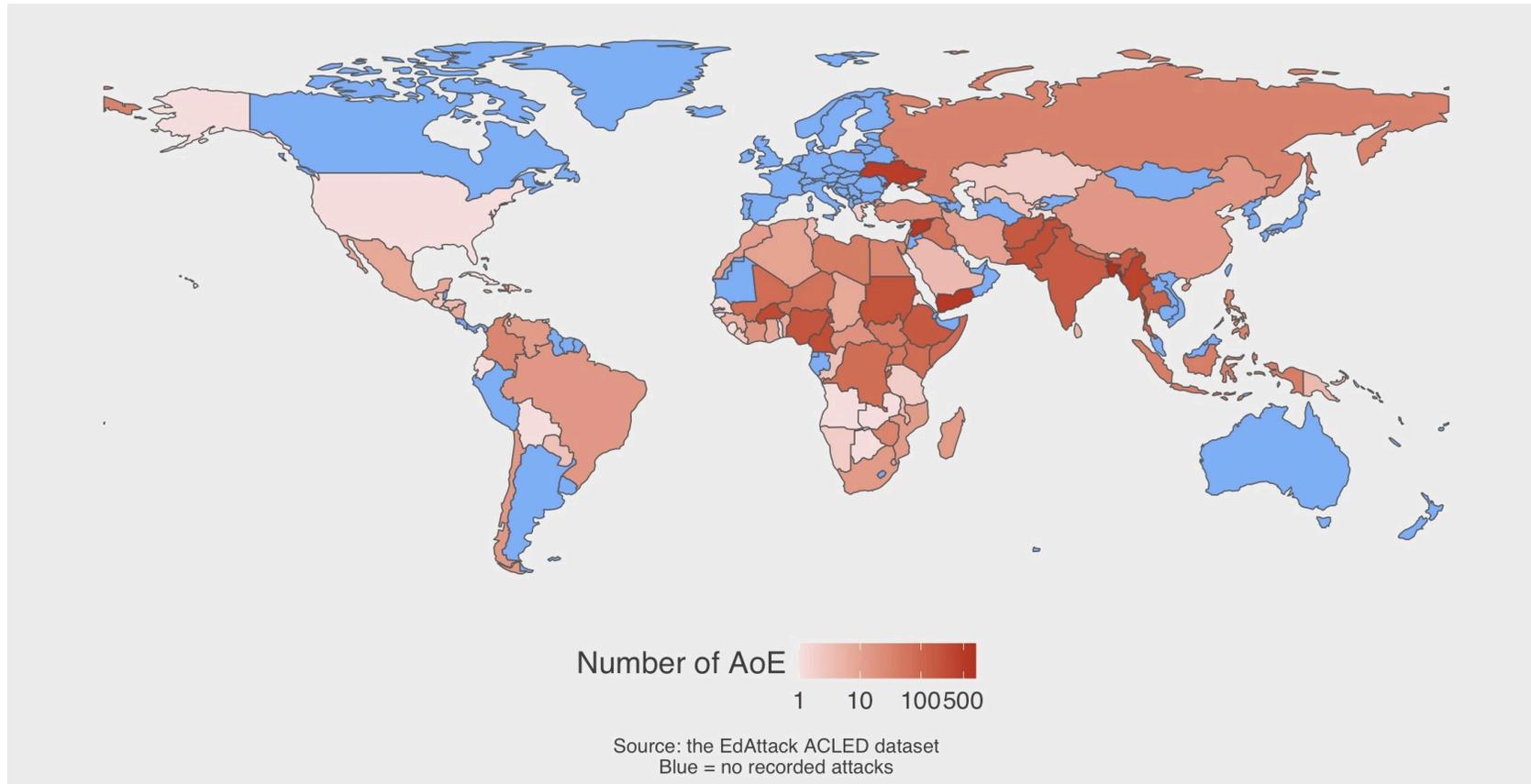
# Test-set Performance

Group	Label	Precision	Recall	F1	N	Pred
<b>Core modality</b>	Violence	0.907	0.862	<b>0.884</b>	283	269
	Arrest & Detention	0.600	0.875	0.712	24	35
	Recruitment & Abduction	0.892	0.967	<b>0.928</b>	60	65
	Sexual Violence	0.909	0.952	<b>0.930</b>	21	22
	Threats/Intimidation	0.750	0.500	0.600	6	4
<b>Target</b>	Student	0.873	0.920	0.896	150	158
	Personnel	0.912	0.929	<b>0.920</b>	112	114
	Infrastructure	0.904	0.904	0.904	125	125
<b>Victim gender</b>	Female mentioned	0.735	0.857	0.791	42	49
	Male mentioned	0.774	0.911	0.837	45	53
<b>Institution</b>	Female-focused	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	17	17
	Male-focused	0.000	0.000	0.000	2	0

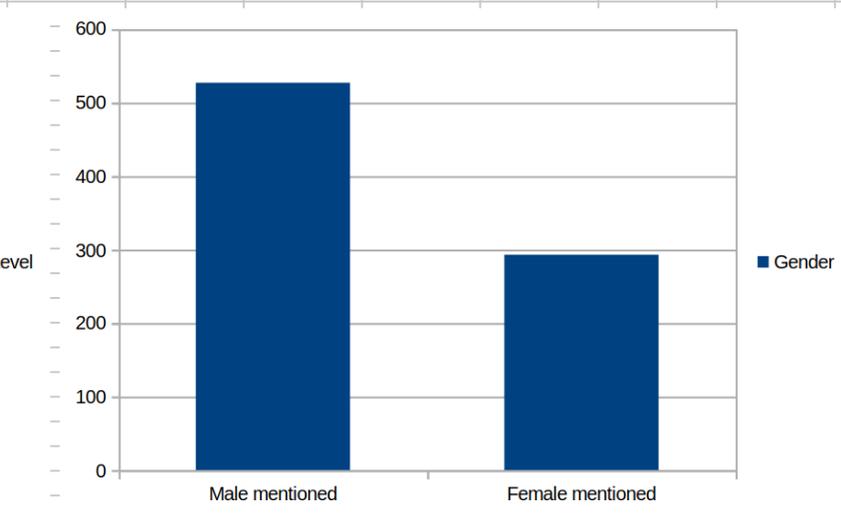
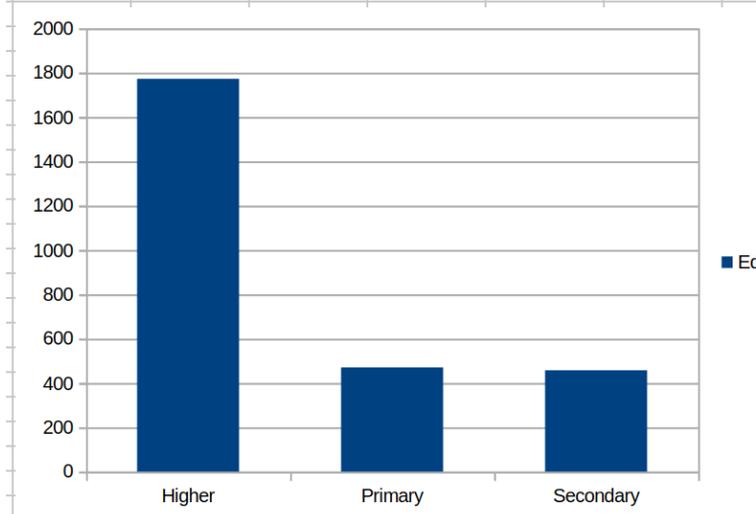
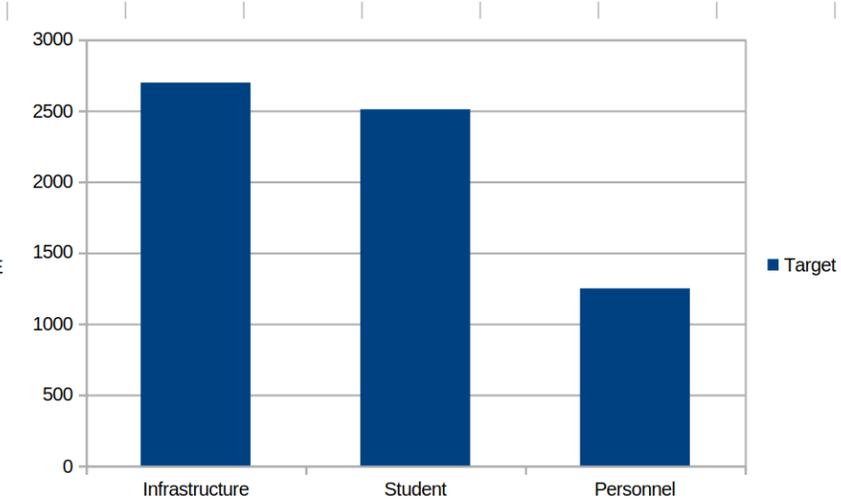
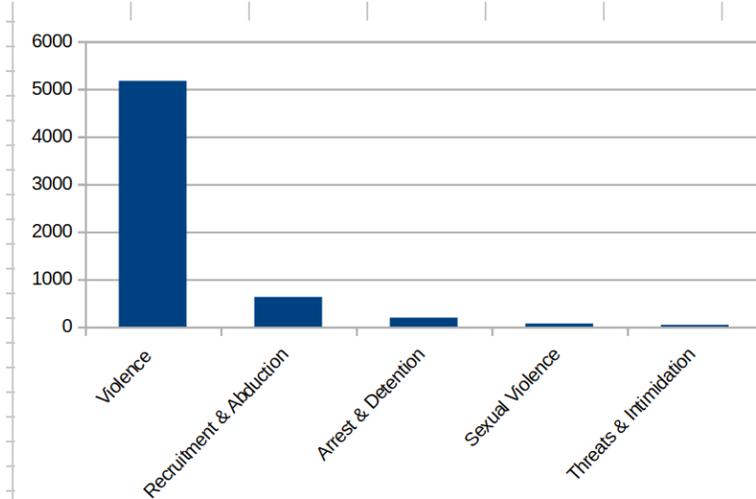
## Test-set Performance

Group	Label	Precision	Recall	F1	N	Pred
Education level	Primary	0.917	0.898	<b>0.907</b>	49	48
	Secondary	0.811	1.000	0.896	43	53
	Higher	0.837	0.885	0.860	87	92
Additional flags	Incidental Flag	0.211	0.200	0.205	20	19
	No attack on education	0.930	0.845	<b>0.886</b>	220	200
	Educational Role Uncertain	0.752	0.831	0.789	124	137
Exact Match Ratio				<b>0.680</b>		
Micro-average				<b>0.866</b>		
Macro-average				<b>0.775</b>		
Weighted average				<b>0.867</b>		

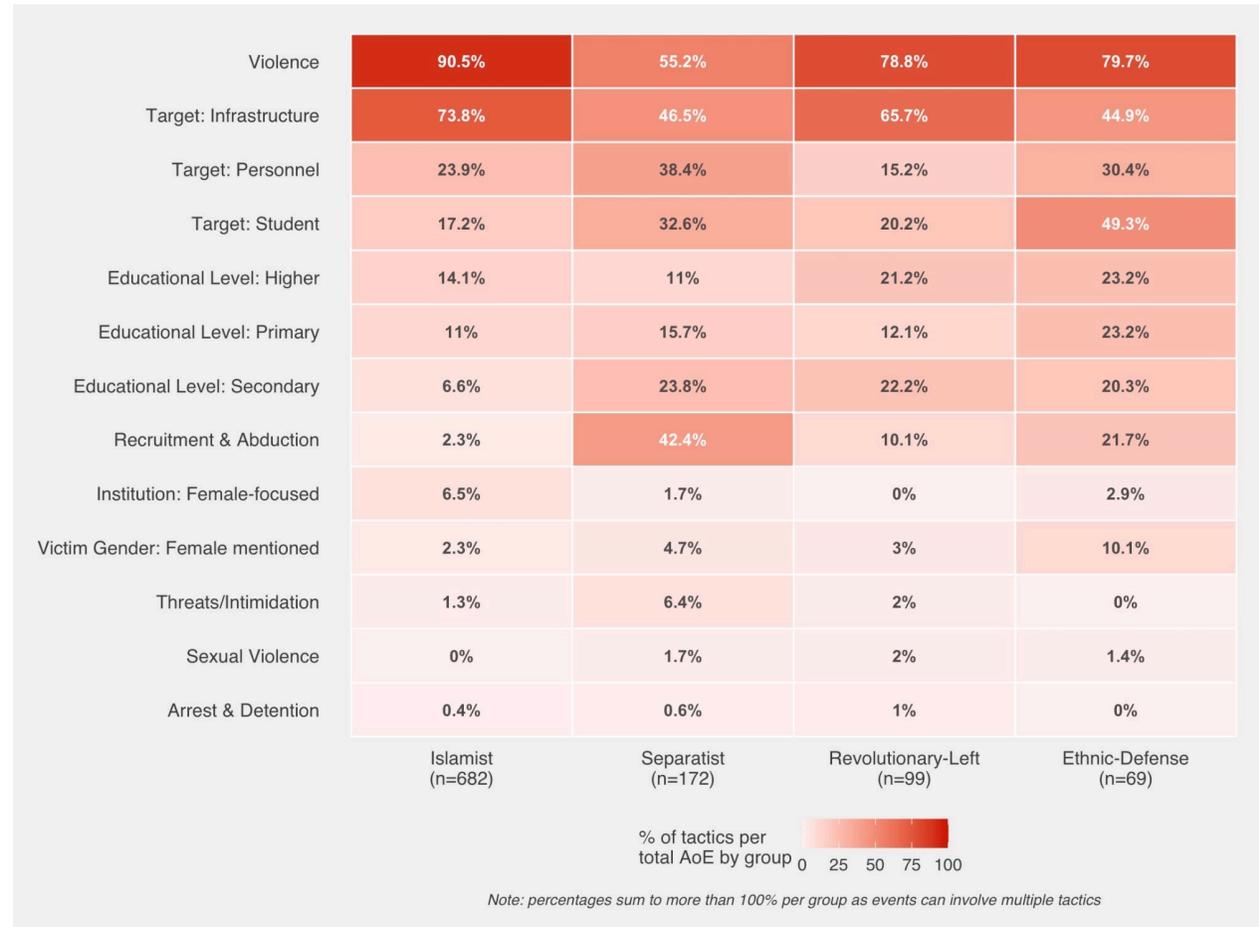
# Attacks on Education Worldwide



# Attacks on Education Worldwide



# Attacks on Education Worldwide



 **Questions?**

