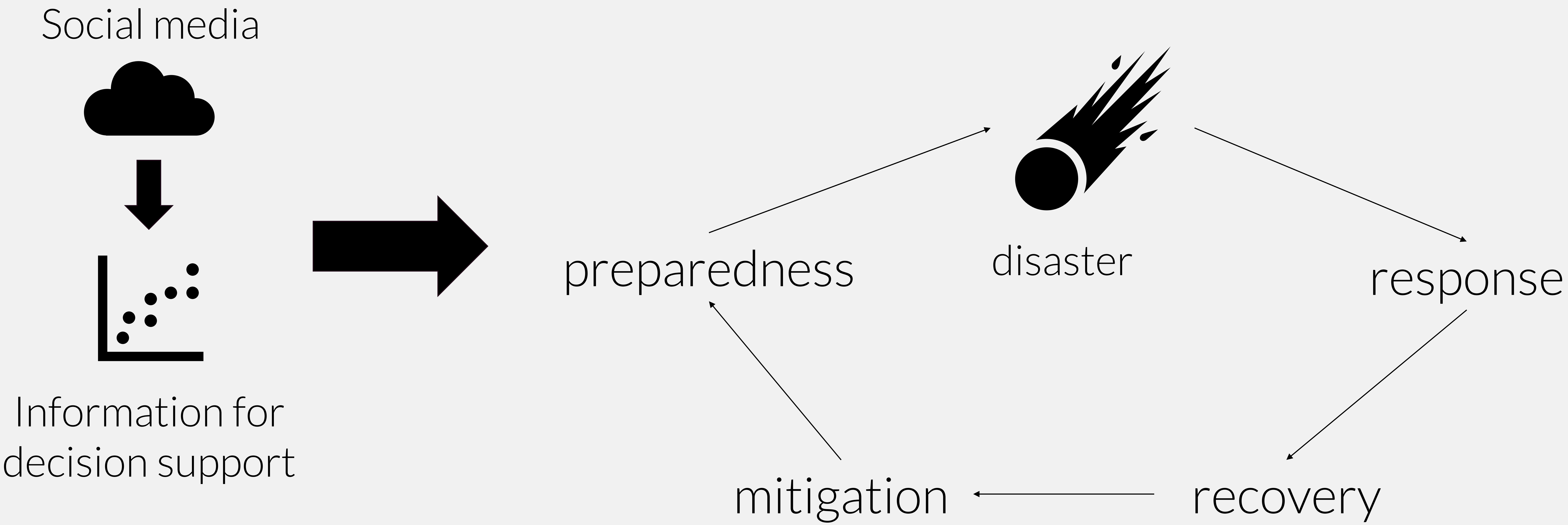
IntelliSys 2024
5-6 September | Amsterdam

Active Learning for Identifying Disaster-Related Tweets
A Comparison with Keyword Filtering
and Generic Fine-Tuning

David Hanny

Decision support using social media

(Geo-)social media as a data source for supporting decision-making in disaster management

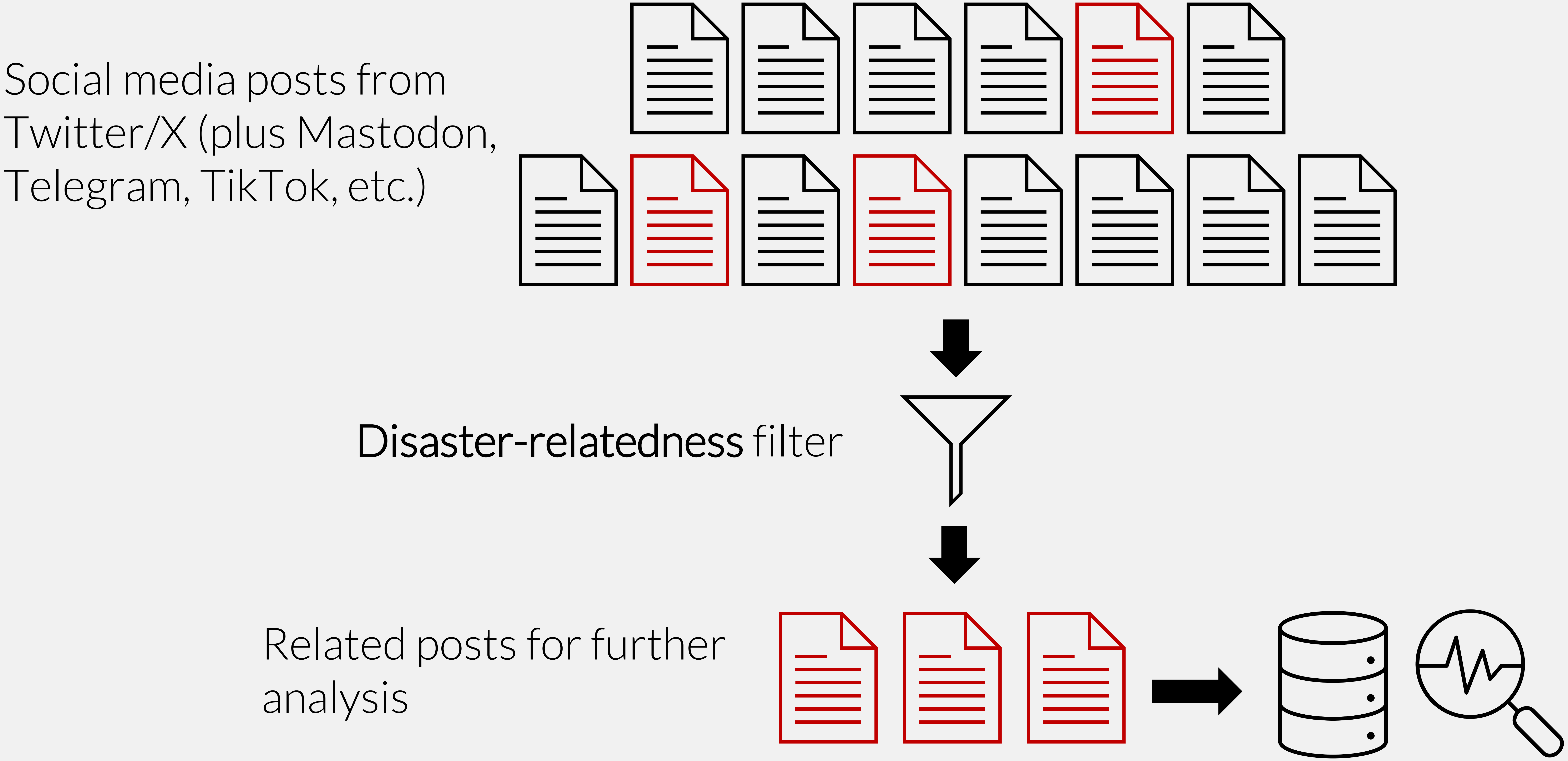


Example (2021 Ahr Valley flood in Germany)
RIP washing machines. One car was trapped inside the underground garage when it started flooding.... 😟 I couldn't find my gummi boots and it was very dark in the basement. 😑 It was a great mistake... http

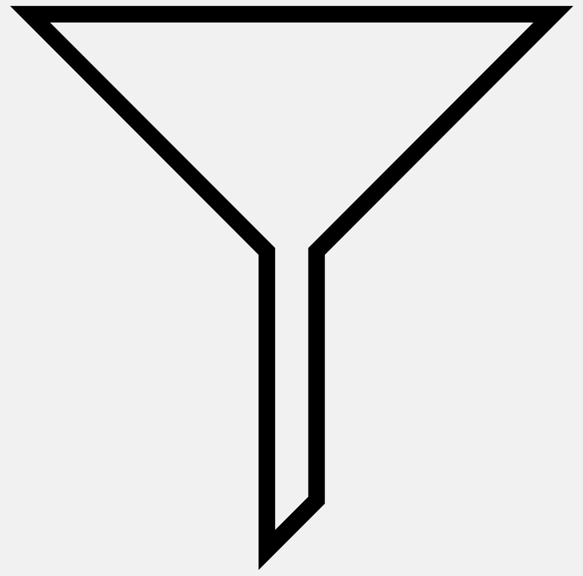
How to filter out the important?

The last few days have been horrible. My home suffered a bad flooding. The water is still high, people are still missing. Right now there’s 50 people confirmed that have died. http
→ disaster-related

Tiny little froggies from a farmer's local pond. 🐸💚 http
→ not disaster-related



Improving the filter – the options

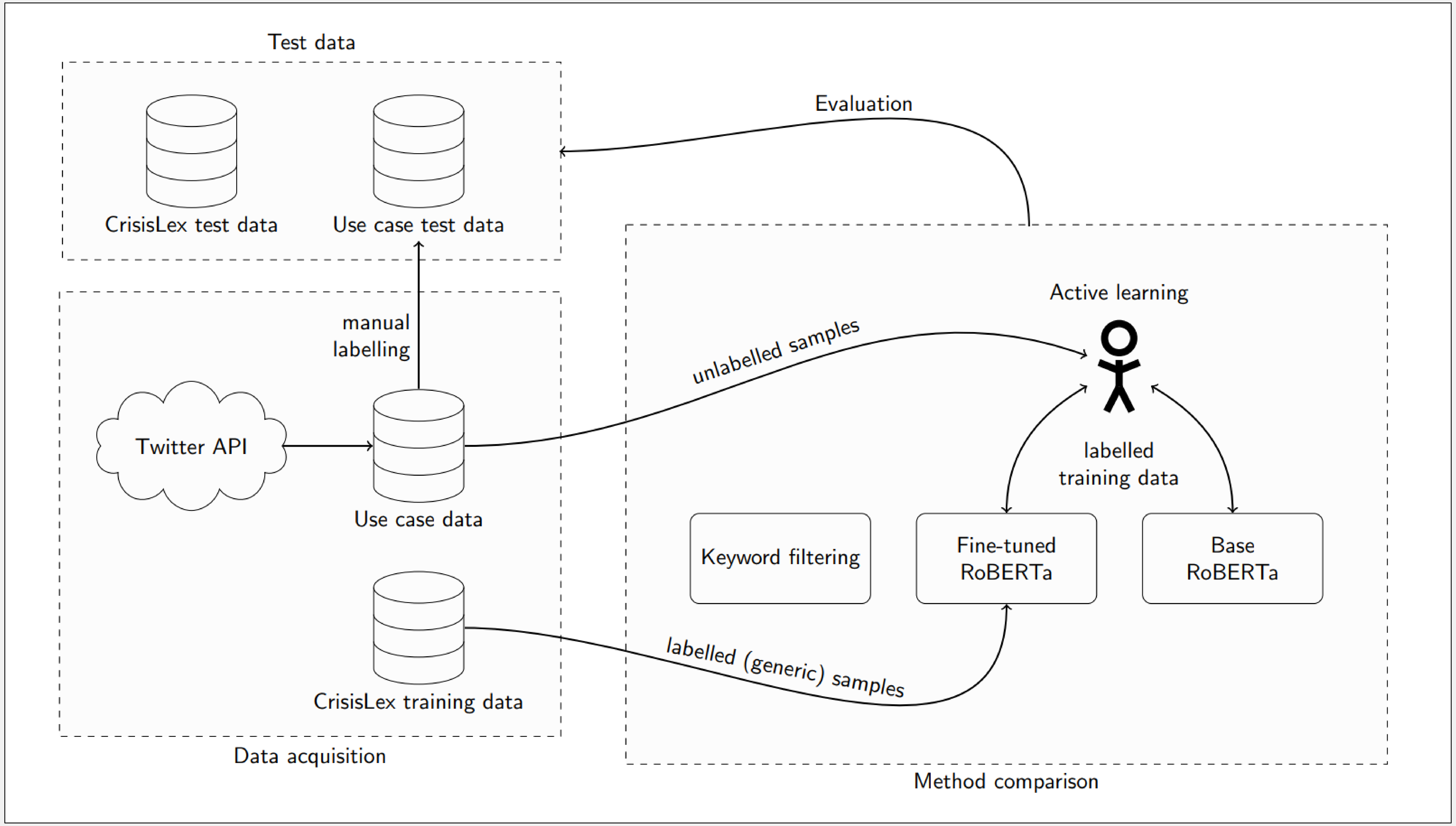


Keyword filtering
- quick and easy
- needs pre-selected keywords or hashtags
- e.g. Shah et al. (2021), Chen et al. (2018)

Supervised Techniques (BERT, RoBERTa, CNNs)
- includes semantic context beyond words
- requires significant training data
- e.g. Madichetty et al. (2023), Koshy et al. (2023)

Active Learning (semi-supervised)
- can significantly reduce needed training data (Settles, 2009)
- rarely used, but examined e.g. by Paul et al. (2023)

Active learning vs others



RQ: How does an AL-based approach compare to keyword filtering or fine-tuning using a broad generic data set for the classification of disaster-related Tweets?

Data: Generic and specific

Generic

Labelled natural disaster tweets from CrisisLexT6 (Olteanu et al. 2015) and CrisisLexT26 (Olteanu et al. 2016)

Label	New Label
on-topic, related and informative, related but not informative	related (1)
Off-topic, not related, not applicable	unrelated (0)

→ Translated to Spanish, German, Italian and French (224,239 tweets)

Use-case specific

Collected via former Twitter API
2021 German Ahr Valley flood (July)
2023 Chile forest fires (January-May)

Use case	#tweets	#labelled (0/1)
2021 Germany flood	11,175	192
2023 Chile forest fires	1,739,986	364

(1) Keyword filtering

Keywords	Languages	Data
earthquake, volcano, landslide, fire, flood, tornado, typhoon, erdbeben, vulkan, erdrutsch, feuer, flut, überschwemmung, wirbelsturm, taifun, terremoto, volcán, deslizamiento, incendio, inundación, tifón, tremblement de terre, volcan, glissement de terrain, incendie, inondation, tornade, typhon	en, de, es, it	CrisisLex
flut, hochwasser, überschwemmung, inundation, flood, disaster, verstroming, hoogwater, vloed, inondation, crue, marée haute	de, en, nl, fr	2021 Germany flood
incendio, forest fire, fuego forestal, bosque quemado	es, en	2023 Chile forest fires

→ Absolute matching and fuzzy matching using the string edit distance (Levenshtein, 1965) with threshold 2.

(2) Fine-tuning with generic data

Fine-tune Twitter-XLM-RoBERTa-base (Barbieri et al. 2022) for binary classification using labelled generic data from CrisisLex.

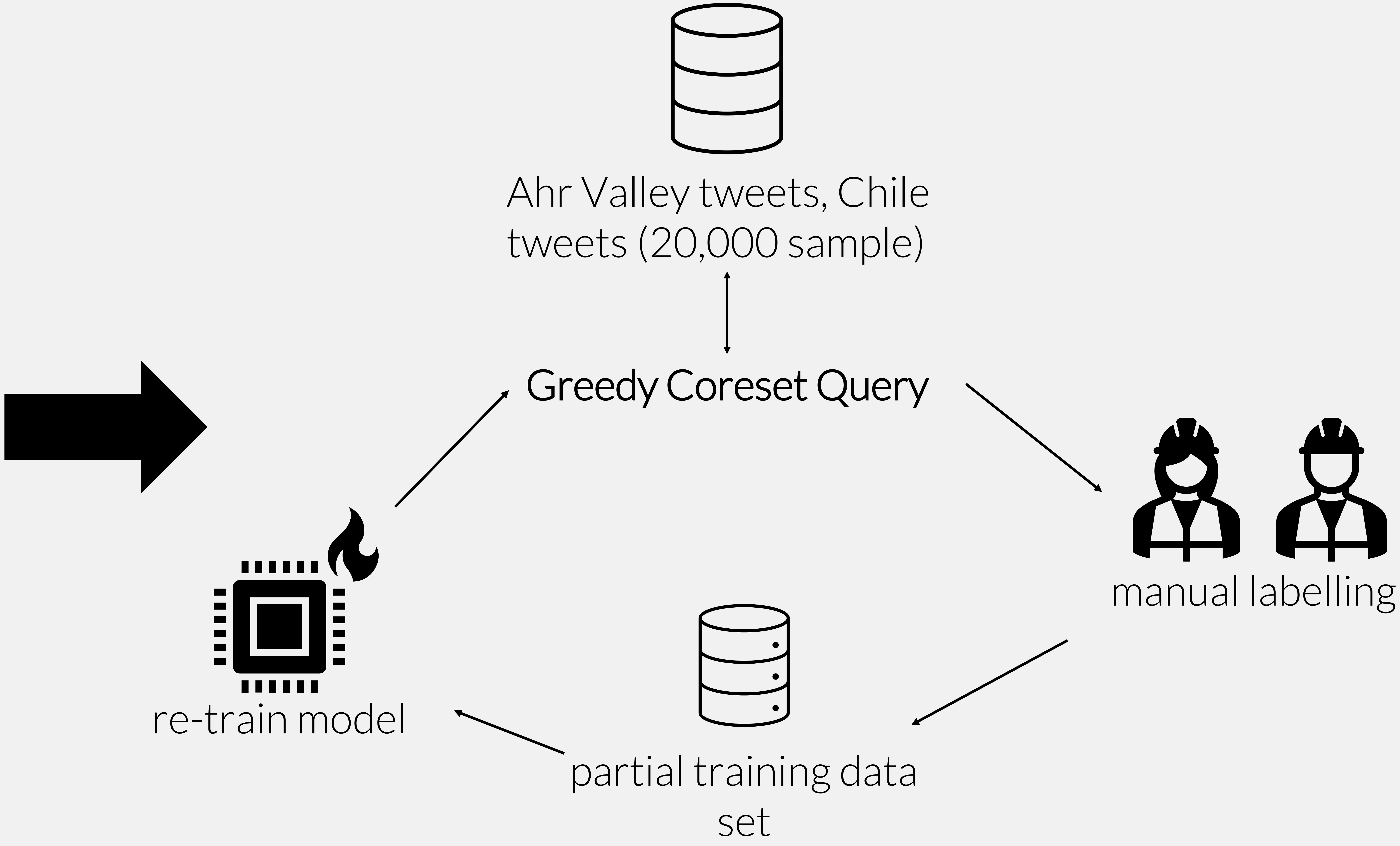
In theory: low effort option

179,391 training data points

(3) Active learning

Untouched Twitter-XLM-RoBERTa-base

Generically fine-tuned Twitter-XLM-RoBERTa-base



Sener, O., & Savarese, S. (2018, February 15). *Active Learning for Convolutional Neural Networks: A Core-Set Approach*. International Conference on Learning Representations.

Results

GFT + AL consistently outperformed all other approaches.
AL only had some dips in recall and precision.

GFT performed well for the generic CrisisLex data but was not as suited for use-case specific data.

KWF yielded mixed results but was better than originally assumed.

Value pairs consist of (”unrelated” (0) / ”related” (1)).

	KWF	Fuzzy KWF	GFT	AL	GFT + AL
(a) Evaluation metrics for CrisisLex					
Precision	0.61 / 0.92	0.64 / 0.85	**0.96** / 0.92	0.53 / **0.95**	0.94 / 0.94
Recall	0.95 / 0.48	0.88 / 0.59	0.90 / **0.97**	**0.98** / 0.27	0.93 / 0.95
F1 score	0.74 / 0.63	0.74 / 0.69	**0.93** / **0.95**	0.69 / 0.41	**0.93** / 0.94
Accuracy	0.70	0.72	**0.94**	0.59	**0.94**
(b) Evaluation metrics for 2021 Germany flood					
Precision	0.96 / **1.00**	0.96 / 0.86	**0.98** / 0.77	0.94 / 0.82	**0.98** / 0.87
Recall	**1.00** / 0.71	0.98 / 0.75	0.96 / **0.83**	0.98 / 0.58	0.98 / **0.83**
F1 score	**0.98** / 0.83	0.97 / 0.80	0.97 / 0.80	0.96 / 0.68	**0.98** / **0.85**
Accuracy	**0.96**	0.95	0.95	0.93	**0.96**
(c) Evaluation metrics for 2023 Chile forest fires					
Precision	0.65 / 0.79	0.65 / 0.79	0.63 / 0.69	0.62 / 0.77	**0.74** / **0.86**
Recall	0.82 / 0.61	0.82 / 0.61	0.67 / 0.65	0.82 / 0.55	**0.87** / **0.73**
F1 score	0.73 / 0.69	0.73 / 0.69	0.65 / 0.67	0.71 / 0.64	**0.80** / **0.79**
Accuracy	0.71	0.71	0.66	0.68	**0.80**

Conclusion and beyond

Learnings

- AL on a pure Twitter-XLM-RoBERTa-base model did not perform all that well.
- AL on top of generic fine-tuning outperformed all other approaches.

Outlook

- Our model provides a basis for future work on geo-social media analysis in disaster management.
- Future work concerns comparing our approach to zero-shot labelling with generative LLMs (e.g. GPT4, Llama-3).

Download the model

Download the model

Hugging Face

Read the paper

TEMA

TRUSTED
EXTREMELY PRECISE
MAPPING AND PREDICTION
FOR EMERGENCY
MANAGEMENT

Funded by the **TEMA project** of the European Commission - European Union under HORIZON EUROPE (HORIZON Research and Innovation Actions), grant agreement 101093003

Email: david.hanny@plus.ac.at
Research group website: https://geosocial.at/