EMNLP 2025 Suzhou, China 中国苏州

FONDAZIONE BRUNO KESSLER

Multilingual *vs* Crosslingual Retrieval of Fact-Checked Claims

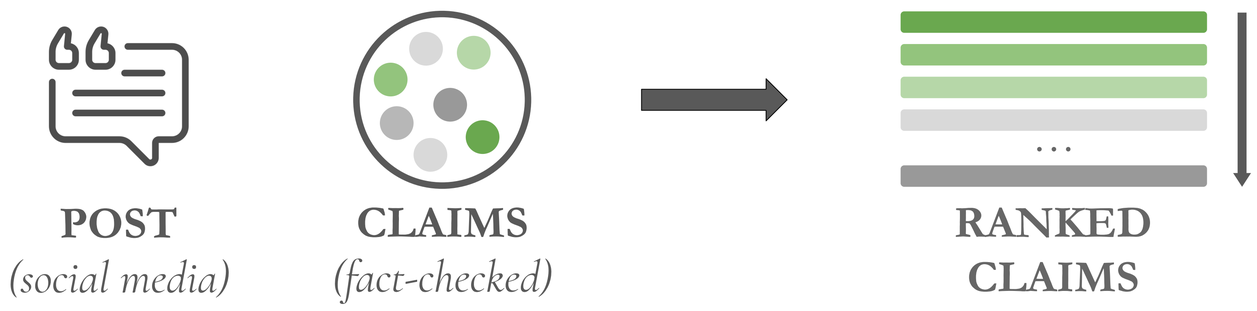
A Tale of Two Approaches

Alan **Ramponi**,[1*] Marco **Rovera**,[1*] Robert **Moro**,[2] Sara **Tonelli**[1]

[1]Fondazione Bruno Kessler [2]Kempelen Institute of Intelligent Technologies [*]*Equal contribution*

TASK

Given a **post** and a set of **previously fact-checked claims**, ***rank*** the claims so that the most relevant ones are ranked as high as possible



DATA

Our subset of the **MultiClaim v2** dataset

- **63k+** *multilingual*, **9k+** *crosslingual* post–fact-checked claim pairs
- **47** languages, for a total of **263** language combinations
- Posts: *Facebook* (45.1k), *Twitter/X* (7.3k), *Instagram* (2.6k), *Telegram* (0.4k)

DATA CODE

METHODS

Unsupervised setting

- Dense **retrieval** and **re-ranking**
  - ***Multilingual TEMs***: 16 models selected among the top-performing ones in the MTEB benchmark
  - ***Cross-encoder re-ranker***: `bge-reranker-v2-m3`
  - ***LLM-based re-ranker***: `RankGPT`
  - ***Claims to re-rank***: top-*n* = 30 claims
- Evaluation on ***full*** set and ***test*** set

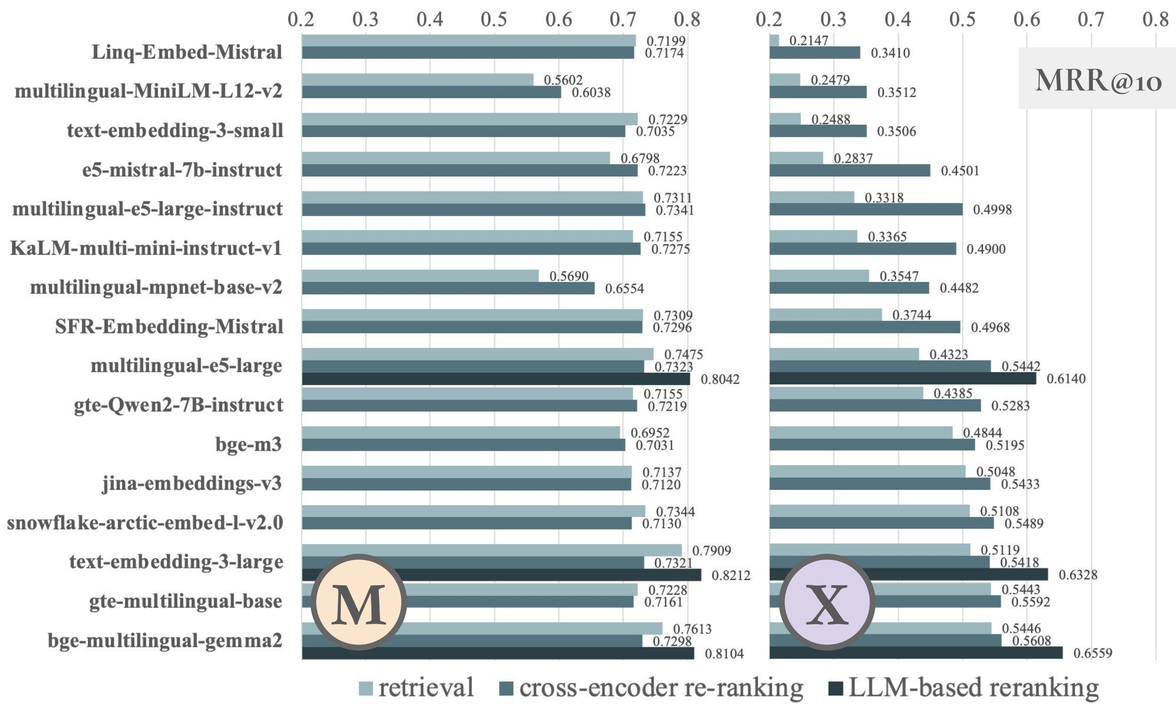
Supervised setting

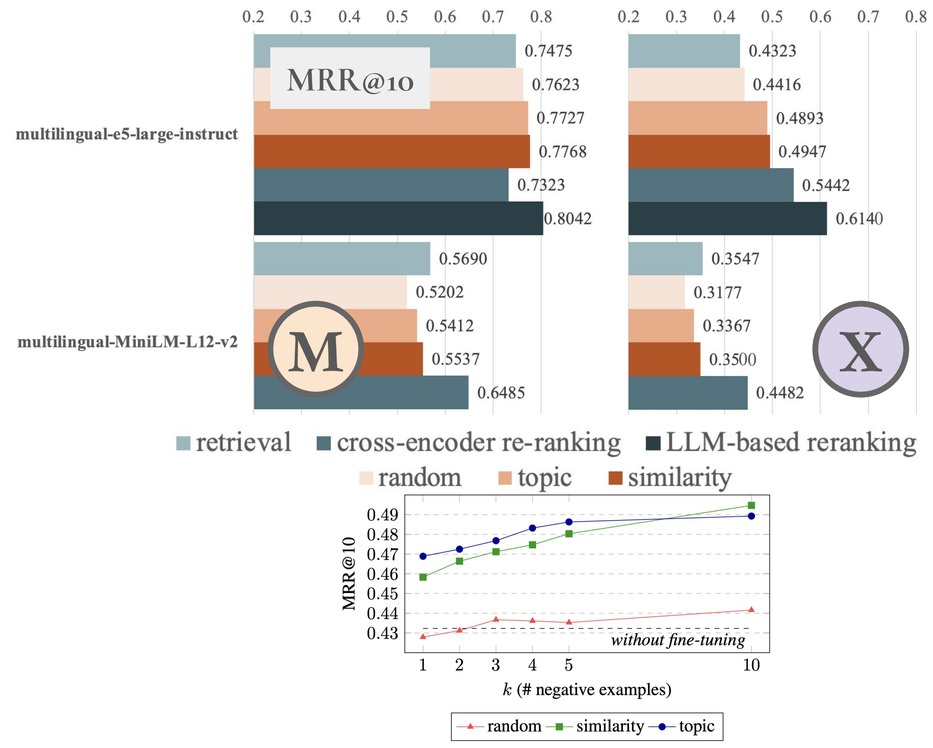
- **Fine-tuning** using **sampled negative** pairs
  - ***Multilingual TEMs***: `multilingual-e5-large` and `multilingual-mpnet-base-v2`
  - ***Sampling strategies***: random, similarity, and topic
  - ***Number of negatives***: 1, 2, 3, 4, 5, 10
  - ***Loss***: multiple negatives ranking
- Evaluation on ***test*** set

Best performing on the dev set

RESULTS & FINDINGS

All methods are evaluated using MRR@10 and S@10 metrics in *multilingual* (M) and *crosslingual* (X) settings





- Models perform **differently** in (M) and (X) settings
- Re-ranking generally effective, **major impact** in (X)
  - *LLM-based* better than *cross-encoder*, but ⬆️ cost
- Model size (*#params*) and embedding dimension **do not correlate** with performance
- *Similarity* strategy better than *topic* & *random* in both (M) and (X) settings (>5 negatives)
- With `multilingual-e5-large`, similarity better than *retrieval* ((M),(X)) and *cross-encoder* in (M), but worse than *LLM-based* and ⬆️ cost