# CS779 COURSE PROJECT

# MULTILINGUAL AND CROSSLINGUAL FACT-CHECKED CLAIM RETRIEVAL


**GROUP - 4**

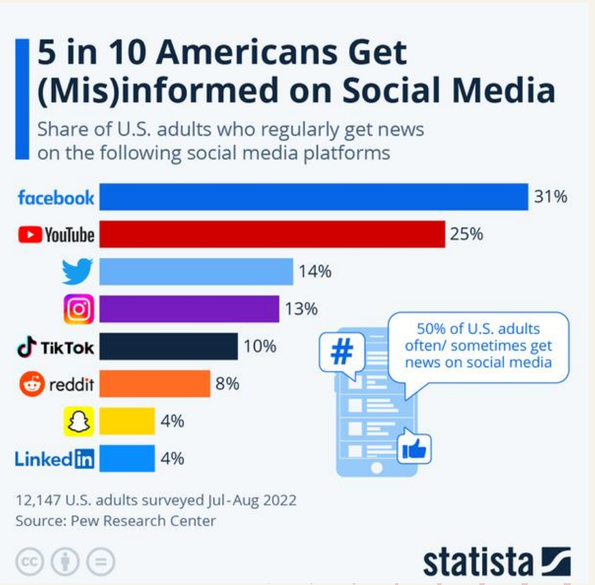
Rahul Jha - 210805

Monika Kumari - 210629

Chirayush Mohanty - 210289

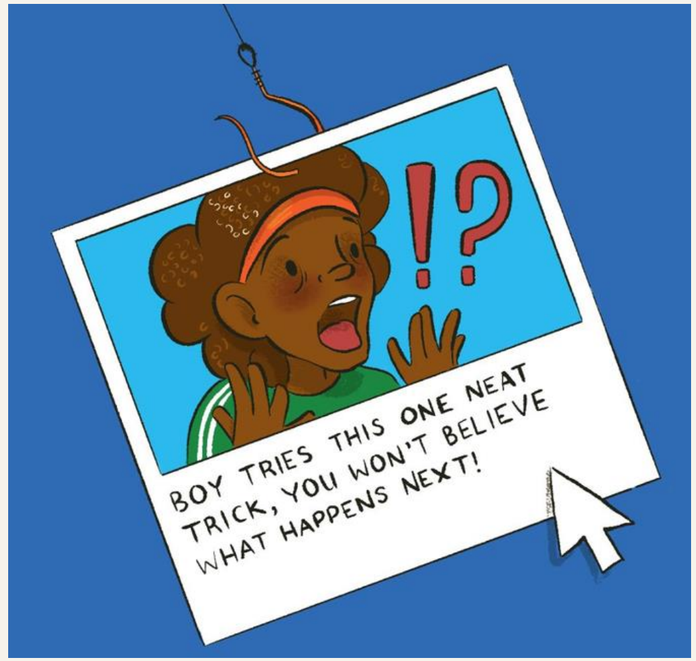
# PROBLEM STATEMENT

- To retrieve relevant fact-checked claims for given social media posts across multiple languages.
    - Addressing the challenge of efficiently identifying previously fact-checked claims in a multilingual and cross-lingual context.
    - Supporting fact-checkers and researchers in their efforts to curb the global spread of misinformation.

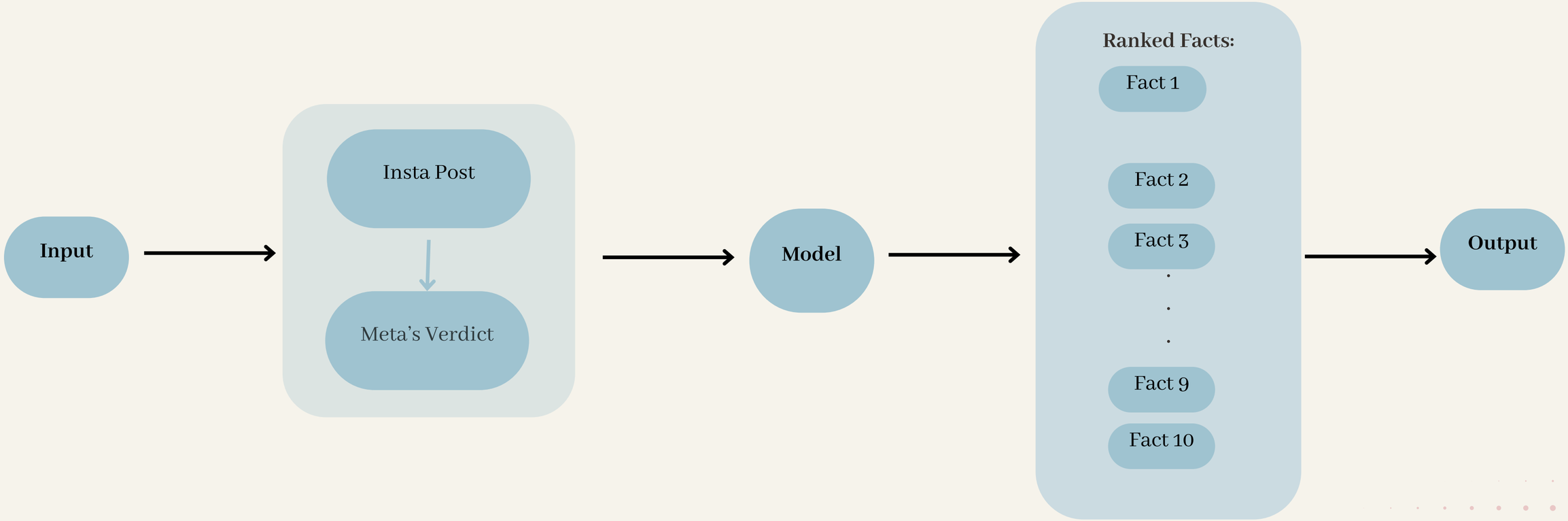


# Example

An Instagram post claims in indonessian “vaksin polio menurunkan memori otak” ( polio vaccines decrease brain memory). Existing fact-checks might be published in English or French, stating, "There is effect of polio vaccines on brain memory." If the fact-checker only speaks indonessian, they might miss crucial information simply because of the language barrier. This inefficiency not only wastes time but also allows misinformation to spread unchecked across different linguistic communities.



# Model Expectation:



# LITERARY REVIEW

1. **"X-Fact: A New Benchmark Dataset for Multilingual Fact-Checking" (Jiang et al., 2021),** achieved an average F1 score of **80.2%** across the 15 languages, with particularly strong results in high-resource languages like English and Spanish.

2. **Multilingual Fact-Checking via  Cross-Lingual Transfer" (Xu et al., 2020),** The paper applied cross-lingual transfer learning using XLM-R, fine-tuning it on fact-check retrieval tasks. The system achieved an accuracy of **83.5%** in retrieving relevant fact-checks for multilingual posts, showing strong performance in cross-lingual tasks.

3. **Fact Extraction and Verification Across Languages" (Thorne et al., 2020)**, The authors used BERT-based retrieval models for fact-checking across languages. They focused on leveraging multilingual embeddings to extract and verify facts from cross-lingual sources. The approach achieved **81.2%** accuracy for retrieving relevant fact-check claims

4 **Cross-Lingual Fact-Checking with Minimal Supervision" (Wu et al., 2021),** The paper proposed a zero-shot learning approach for fact-checking. It finetunes mBERT to handle multilingual inputs using unsupervised data augmentation techniques for fact-check retrieval. Model achieved an accuracy of **79.4%**

5 **Multilingual Evidence Retrieval for Fact-Checking" (Zhou et al., 2022),** The authors used a dual-encoder model with contrastive loss, to learn multilingual embeddings for fact-check retrieval. They used mBERT to encode both claims and evidence. The vsystem achieved an F1 score of **82.7%**

In terms of evaluation, **Success-at-K (S@K)** was the primary metric used, with results showing that **GTR-T5** outperformed other models in both monolingual and cross lingual tasks, especially when coupled with machine translation. Interestingly, larger models did not always yield better results, as architecture and training techniques proved to be more crucial. The inclusion of visual content through OCR also played a role in enhancing the retrieval process, although it was noted that discrepancies due to missing visual context affected performance.

**XLM-R (Conneau et al., 2020)** is a **State-of-the-Art (SOTA)** model for many **crosslingual and multilingual tasks**, such as **XNLI**, **MLQA**, **TyDiQA**, and **XTREME**.

| | Reader | Retrieval Method | Zero-shot F1 (%) | In-domain F1 (%) |
|---|---|---|---|---|
| Prior (Gupta and Srikumar, 2021) | Majority | None | 7.6 | 6.9 |
| | mBERT | None | 16.7 | 39.4 |
| | mBERT | Google Search | 16.0 | 41.9 |
| Ours | mBERT | None | 17.25 | 36.91 |
| | mBERT | Google Search | 16.02 | **42.61** |
| | mBERT | MT+DPR | 15.01 | 35.29 |
| | mBERT | BM25 | 17.43 | 38.29 |
| | mBERT | mDPR | 17.60 | 36.79 |
| | mBERT | CONCRETE | **19.83*** | 40.53 |

source: https://aclanthology.org/2022.coling-1.86.pdf

| Split | # claims | # languages |
|---|---|---|
| Train | 19079 | 13 |
| Development | 2535 | 12 |
| In-domain | 3826 | 12 |
| Out-of-domain | 2368 | 4 |
| Zero-shot | 3381 | 12 |

Table 1: Dataset statistics of X-FACT.

source:https://aclanthology.org/2021.acl-short.86.pdf

# DATASET



- **The dataset provided for the project consists of two main sources: fact-check data and social media posts, alongside a mapping between the two.**
- **The fact-checks are sourced from various fact-checking websites and the social media posts span multiple platforms.**
- **The data is multilingual, covering 27 languages, and is multimodal in nature, incorporating text extracted from images via OCR.**

## Fact-Check Data (fact_checks.csv)

- **fact_check_id:** Unique identifier for each fact check.
- **claim:** Original claim, its translated version, and the language.
- **instances:** List of timestamps and URLs where the fact check is mentioned.
- **title:** Original title, its translated version, and the language.

## Mapping (pairs.csv)

- **fact_check_id:** Links to fact_checks.csv.
- **post_id:** Links to posts.csv.

## Social Media Posts (posts.csv)

- **post_id:** Unique identifier for each social media post.
- **instances:** List of timestamps and social media platforms where the post appears.
- **ocr:** Text extracted from images in the post, along with translations.
- **verdicts:** Labels attached by platforms (e.g., "False information").
- **text:** The main text of the post and its translated version.

*fact_checks.csv : 153743 entries*

| fact_check_id | claim | instances | title |
|---|---|---|---|
| 12 | ('!! Ø¨ÙŠØ´Ø±Ø¨ Ù…ÙŠØ§Ù‡ Ù…Ø¹Ø¯Ù†ÙŠØ© Ù. | [(None, 'https://dabegad.c | ('Ø-Ù,ÙŠÙ,Ø©Ø´Ø±Ø¨ Ø§Ù„Ø³Ù |
| 13 | ('" As vacinas nÃ£o passaram pelos protocolos de te | [(1614511874.0, 'https://c | ('Fact Check. As vacinas contra a |

*posts.csv : 24437 entries*

| post_id | instances | ocr | verdicts | text |
|---|---|---|---|---|
| 2751 | [(1604555943.0, 'fb')] | [('t YEARNING - HOME F | ['False information'] | ('!Boletas tiradas! |
| 2752 | [(1604588844.0, 'fb')] | [('UNITED STATES POST | ['False information'] | ('!Boletas tiradas! |

*pairs.csv : 25743 entries*

| post_id | fact_check_id |
|---|---|
| 2228 | 33 |
| 2228 | 23568 |

# VERDICTS FROM META VS NO. OF POSTS

| | | | |
|---|---|---|---|
| 1. | False information | 12408 | |
| 2. | Partly false information | 3410 | |
| 3. | Missing context | 1355 | |
| 4. | False information. | 1147 | |
| 5. | Altered photo | 493 | |
| 6. | Partly false information. | 351 | |
| 7. | Partly False | 271 | |
| 8. | Missing context. | 256 | |
| 9. | False | 241 | |
| 10. | False information and graphic content | 222 | |
| 11. | Altered video | 108 | |
| 12. | Missing Context | 88 | |
| 13. | Sensitive content | 33 | |
| 14. | Altered photo/video. | 15 | |
| 15. | Altered Photo/Video | 14 | |
| 16. | False headline | 2 | |
| 17. | Support your streamers by sending them stars | 2 | |
| 18. | Altered photo/video | 1 | |

# PREDICTION

- A single JSON file where each key is a post_id (as a string) and the corresponding value is a list of up to 10 fact_check_ids (as integers) that are most relevant to that post

- Key: post_id (as a string).
- Value: A list of 10 fact_check_ids (as integers) that your system predicts to be the most relevant fact checks for that post.

```
{
  "0": [5, 10, 2, 65, 15, 255, 11, 8, 420, 502],
  "1": [12, 13, 5, 0, 125, 450, 220, 18, 49, 51],
  "2": [444, 4, 7, 18, 29, 55, 263, 178, 99, 82],
  ...
  "3840": [11, 3, 507, 624, 177, 39, 20, 66, 344, 327]
}
```

# PRE PROCESSING



- The dataset cleaning process involved tokenizing and removing punctuation and stop words from the text fields (i.e., OCR, title, text, and claims) to ensure uniformity and reduce noise.
- For multilingual claims, both the original and translated versions were cleaned separately.
- Additionally, timestamps and instances associated with posts and fact-checks were processed to better align the temporal context between the two.
- The processed data was prepared for model training by ensuring clean and structured text, removing discrepancies, and creating separate fields for key information like language and timestamp. This approach enabled better alignment and comparison of posts and fact-checks, improving retrieval performance.

# INITIAL MODEL



- Our training model takes text from the post , claims from the facts and labels as input.
- We assigns labels to the text and claims based on given mapping file.
- Then we sample positive (label 1) and negative samples (label 0) and keep negative sample count as three times of positive samples.
- Then we train our model using mBERT and contrastive loss.
- For testing we just gave text and claims as input and model predicted labels
- For now we trained the model on only a small part of our dataset due to lack of RAM and GPU.
- Our accuracy is 70% and precision ~65% and recall was ~68%  and F1-Score was~68%
- Link to collab file

# FUTURE DIRECTIONS

## PIPELINE:

### Preprocessing

- Posts and fact-checks will be tokenized, and paired data from pairs.csv will be used to fine-tune mBERT.
- Optimal no. of negative samples will be generated by random sampling to provide contrasting examples for the model.

### Input/Output

- The new input will consist of the merged text and OCR fields from social media posts, and the merged claims and titles from fact-checks.
- The output will be a ranked list of top-10 fact-check IDs for each post, providing the most likely fact-checked claims that match the content of the post.

### Model Architecture:

- We will fine-tune mBERT using the pairs.csv file for binary classification to distinguish relevant from irrelevant pairs.
- A ranking loss will be applied to ensure the model retrieves the top-10 fact checks that are most relevant to each post.
- Additionally, we will incorporate the instances column, which contains timestamps, and assign higher weights to fact-checks that are more recent compared to the post.
- This approach will allow newer fact-checks, even if slightly less relevant, to be prioritized over older ones, as more recent facts tend to be more accurate in the dynamic social media environment.

# FUTURE TIMELINE

- **Phase 1 - (30 Sept - 10 Oct)**

  Baseline Model Implementation and submission to check benchmark  performance on given dataset

- **Phase 2 - (11 Oct - 25 Oct)**

  Improvent on the model architecture and explore scopes for better performing models for monolinhual fact checks

- **Phase 3 - (26 Oct - 10 Nov)**

  Fine tuning the model and prepare project report and presentation

# INDIVIDUAL CONTRIBUTION

| Name | Contribution |
| --- | --- |
| Rahul Jha | Data Cleaning |
| Monika Kumari | Pipeline code |
| Chirayush Mohanty | Pipeline code |

# CONCLUSION

This project proposes a solution for crosslingual fact-check retrieval using mBERT, addressing a critical problem in misinformation detection. By fine-tuning the model on paired data and applying a ranking loss, we aim to improve the retrieval accuracy of fact-checks across languages. Future work will focus on optimizing the architecture and incorporating additional datasets for fine-tuning.

# THANK YOU