AI for Social Good
Fighting Against Fake News

Hackathon hosted by ARIES - IITD

16-17 March 2024

1 Introduction

We are thrilled to welcome you to ARIES-IITD's first hackathon, focused on **AI for Social Good**, specifically tailored to address the urgent challenge of **combating fake news**.

In an era where misinformation spreads faster than ever, the need for innovative and effective solutions is paramount. Our mission is to harness the power of Artificial Intelligence to create tools that can differentiate between fact and fiction, contributing positively to the discourse on social media and beyond.

Countering fake news is an essential social service for several reasons:

- **Promotes Informed Decision-Making**: Individuals and Organizations help ensure that the public has access to accurate and reliable information, enabling them to make informed decisions.
- **Protects Public Health**: Ensures that people receive accurate guidance on preventive measures, treatments, and vaccine information.
- **Safeguards Democracy**: The dissemination of false information can undermine the democratic process by influencing elections and public opinion through misinformation. Hence, it becomes essential to combat fake news.

The problem statement has been designed with everybody in mind. Everyone from advanced Deep Learning practitioners to ML beginners can attempt this problem statement and perform well. The Hackathon aims to raise awareness about the problem of fake news and data-driven ways to handle it. We hope you find the problem statement fun to solve and enriching at the same time.

2 Problem Statement

The problem statement for this Hackathon revolves around developing Machine Learning models capable of detecting and identifying claims in social media posts. It is hosted on Kaggle. Participants will be working on two critical tasks:

- **Claim Detection**: The first task involves binary classification, where participants will determine whether a given piece of text from a social media post is making a claim or not. This task is essential for filtering noise and focusing on potentially misleading information.

The dataset provided for this challenge consists of a train and a test set. The train and dev set are **.csv** files with **”tweet_text”** and **”claims”(0/1)** values indicating whether a tweet is a claim or not. The test set only consists of tweets.

The task is to predict ’claim’(0/1) given a tweet.

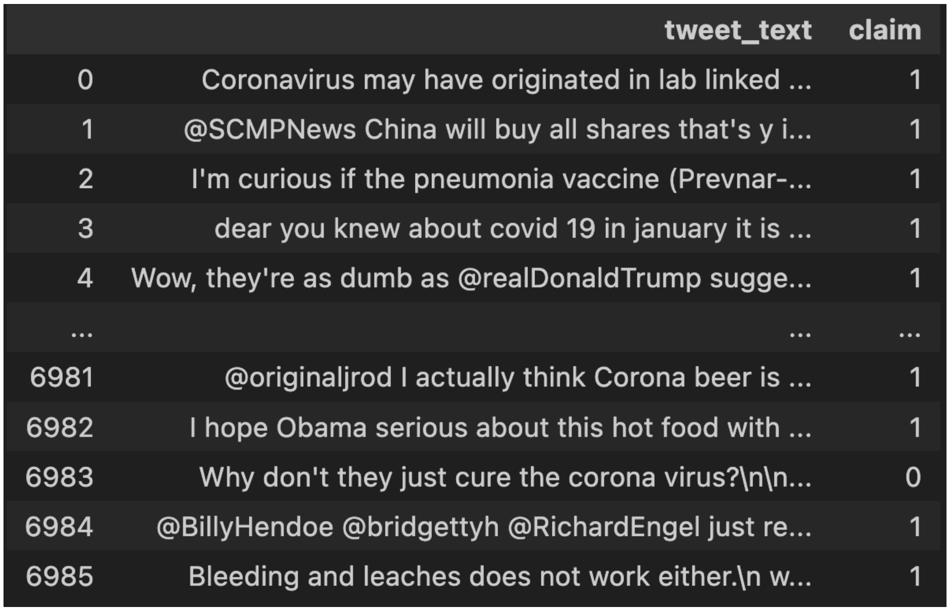
- **Claim Span Identification**: It is a critical task in the process of fact-checking and misinformation analysis, focusing on precisely locating the specific segment of text that constitutes a claim within a larger body of text.

  This task is essential for accurately parsing and understanding the context of a statement, distinguishing it from surrounding non-claim text, and preparing it for further analysis or verification.

  The dataset consists of claims (split into tokens), starting indices of claim spans, and ending indices of claim spans. **span_start_index** consists of a list of integers denoting the start of a claim span. Similarly, **span_end_index** consists of a list of integers denoting the end of a claim span.

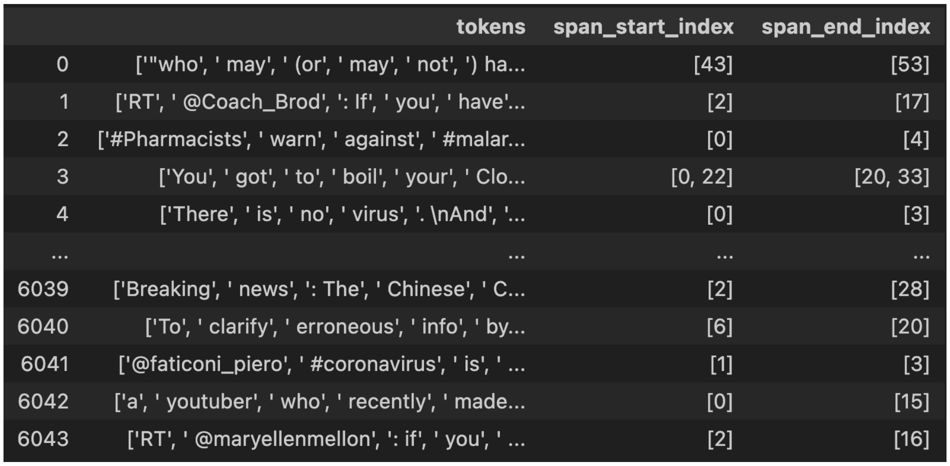
  The task is to provide a list of span_start_index and span_end_index

Claim: A claim is a statement or assertion that presents information as being true, factual, or at least worthy of acceptance or belief. Claims are often made to express an opinion, convey a fact, or argue a point.
Examples:
”The Great Wall of China is visible from space.” : Factual
”Vaccines cause autism.” : Fake



	tweet_text	claim
0	Coronavirus may have originated in lab linked ...	1
1	@SCMPNews China will buy all shares that's y i...	1
2	I'm curious if the pneumonia vaccine (Prevnar-...	1
3	dear you knew about covid 19 in january it is ...	1
4	Wow, they're as dumb as @realDonaldTrump sugge...	1
...	...	...
6981	@originaljrod I actually think Corona beer is ...	1
6982	I hope Obama serious about this hot food with ...	1
6983	Why don't they just cure the corona virus?\n\n...	0
6984	@BillyHendoe @bridgettyh @RichardEngel just re...	1
6985	Bleeding and leaches does not work either.\n w...	1

Figure 1: Training Data for Task 1



	tokens	span_start_index	span_end_index
0	['"who', ' may', ' (or', ' may', ' not', ') ha...	[43]	[53]
1	['RT', ' @Coach_Brod', ': If', ' you', ' have'...	[2]	[17]
2	['#Pharmacists', ' warn', ' against', ' #malar...	[0]	[4]
3	['You', ' got', ' to', ' boil', ' your', ' Clo...	[0, 22]	[20, 33]
4	['There', ' is', ' no', ' virus', '. \nAnd', '...	[0]	[3]
...	...	...	...
6039	['Breaking', ' news', ': The', ' Chinese', ' C...	[2]	[28]
6040	['To', ' clarify', ' erroneous', ' info', ' by...	[6]	[20]
6041	['@faticoni_piero', ' #coronavirus', ' is', ' ...	[1]	[3]
6042	['a', ' youtuber', ' who', ' recently', ' made...	[0]	[15]
6043	['RT', ' @maryellenmellon', ': if', ' you', ' ...	[2]	[16]

Figure 2: Training Data for Task 2

3 Submission Instructions

3.1 Code and Report

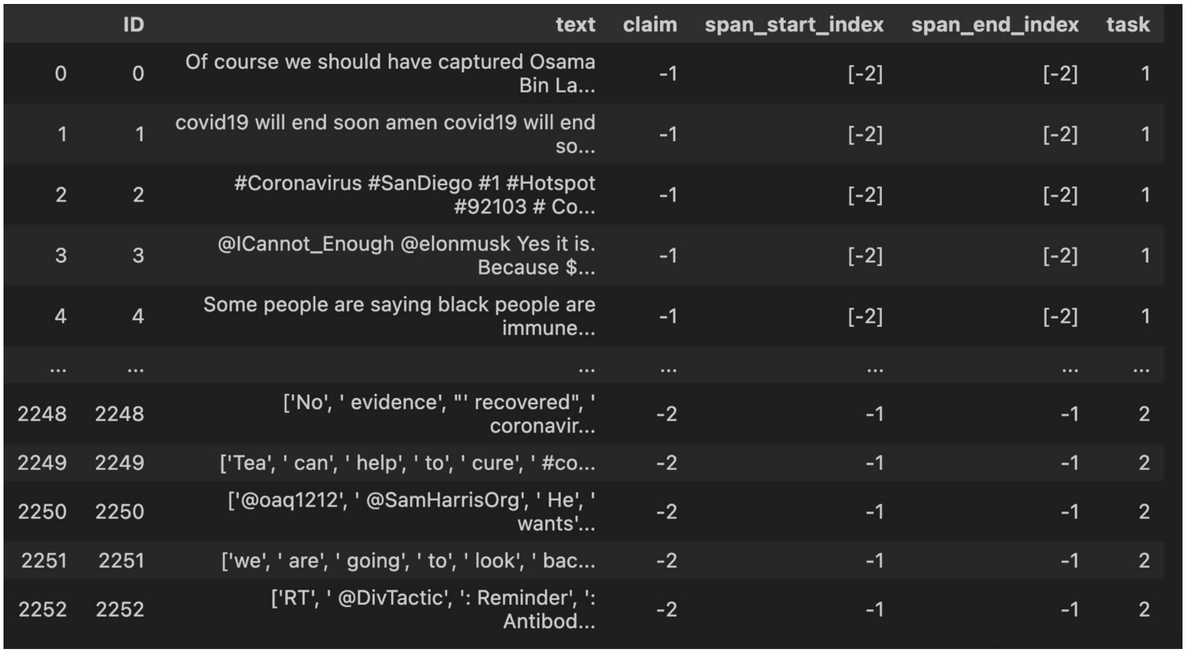
- You are required to submit your code as .ipynb or .py. We will run your implementations to reproduce the results and any discrepancy will results in a disqualification.
- We expect you to submit a 1-2 page report documenting your results, observations and methods.
- Mention in the report what training setup you used (GPU, Epochs, Batch Size etc.)

3.2 Kaggle Submission

On Kaggle, you will be submitting a **.csv** file containing your predictions. The format for the .csv file has been described below and in the ’submission_format.csv’ file.

Due to limitations on submission format on Kaggle, please follow the below rules while submitting your predictions:

- In the following figure there are 6 coulumns, 3 of which are ID, Text and Task, which are fixed.
- In each task (1 & 2) you need to submit your solution for the column(s) with entries -1. The other column(s) with entry -2 will not be evaluated in that task
- For example in task 1, you need to generate only claim.
- **Note :** If you do not intend to submit any of task, please leave the result column values to -1 or -2, whatever it is by default. Failure to following this condition, may lead to errors in producing the score.



	ID	text	claim	span_start_index	span_end_index	task
0	0	Of course we should have captured Osama Bin La...	-1	[-2]	[-2]	1
1	1	covid19 will end soon amen covid19 will end so...	-1	[-2]	[-2]	1
2	2	#Coronavirus #SanDiego #1 #Hotspot #92103 # Co...	-1	[-2]	[-2]	1
3	3	@ICannot_Enough @elonmusk Yes it is. Because $...	-1	[-2]	[-2]	1
4	4	Some people are saying black people are immune...	-1	[-2]	[-2]	1
...	...	...	...	...	...	...
2248	2248	['No', ' evidence', "' recovered", ' coronavir...	-2	-1	-1	2
2249	2249	['Tea', ' can', ' help', ' to', ' cure', ' #co...	-2	-1	-1	2
2250	2250	['@oaq1212', ' @SamHarrisOrg', ' He', ' wants'...	-2	-1	-1	2
2251	2251	['we', ' are', ' going', ' to', ' look', ' bac...	-2	-1	-1	2
2252	2252	['RT', ' @DivTactic', ': Reminder', ': Antibod...	-2	-1	-1	2

Figure 3: Submission Format

4 Evaluation

4.1 What is F1 Score ?

In machine learning, the F1 score is a measure of a model's accuracy, combining both precision and recall. It is the harmonic mean of precision and recall, calculated as:

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

where:

- **Precision** is the number of true positive results divided by the number of all positive results returned by the classifier.
- **Recall** is the number of true positive results divided by the number of all relevant samples (all samples that should have been identified as positive).

The F1 score is a useful metric when you want to balance precision and recall, especially when classes are imbalanced. A high F1 score indicates both good precision and recall, with a perfect F1 score of 1 indicating perfect precision and recall.

4.2 Task 1 Evaluation

- Since the task 1 has only binary classification task, your score is calculated as F1 score of your prediction with respect to the true labels.

4.3 Task 2 Evaluation

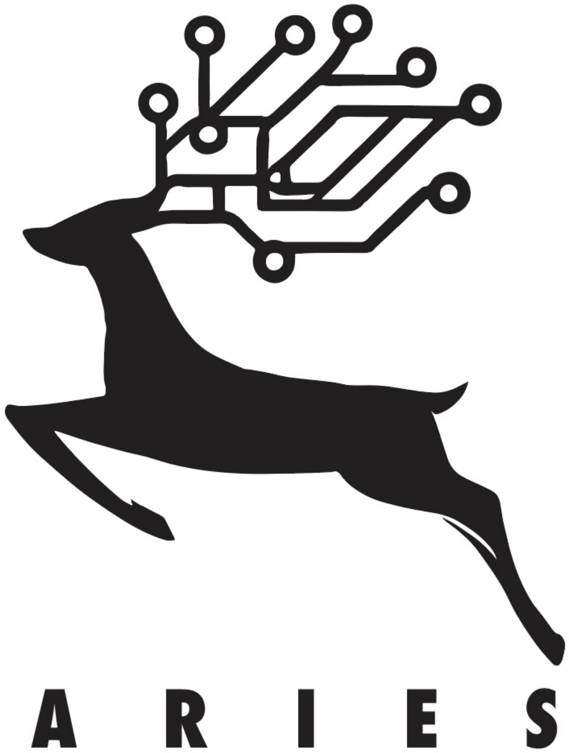
- Token-wise F1-score for each sentence will be calculated, and then an average F1-score over all sentences will be taken.

4.4 Final Score

- The final score will be a weighted average of the two F1-scores. The weights will be revealed towards the end of the competition.

5 Rules and Guidelines

- You must submit your code in .ipynb or .py format. We will execute your code to verify the results, and any discrepancies may result in disqualification.
- You must only use GPUs available for free on online platforms(Kaggle/Colab). This is to ensure that nobody has an undue advantage by having access to a powerful GPU.
- Clearly mention in the report what resources you used.
- Plagiarism will result in disqualification from the event, and a strict ban from future ARIES competitions.
- Use of any fine-tuned model for this task is strictly prohibited. Failure to comply will result in a disqualification from the event.
- If you use anybody else's work or publication, please cite it.
- There is a daily limit of ten submissions. Use them wisely.
- The competition will end on 17th March, at 4:30 PM IST.



@ARIES.IITD