



Team Empty String

Bias Correction With EMBR →

Group Members

Goh Zhen Rong, Tristan Amadeus Surya, Yong Rui Jie

On with the show! ➤

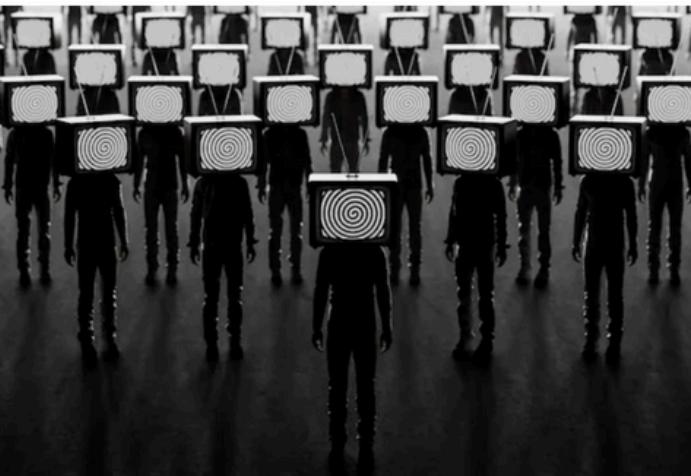


The rise of AI fake news is creating a ‘misinformation superspread’

AI is making it easy for anyone to create propaganda outlets, producing content that can be hard to differentiate from real news

December 17, 2023

6 min



(iStock)

By [Pranshu Verma](#)

Artificial intelligence is automating the creation of fake news, spurring an explosion of web content mimicking factual articles that instead disseminates false information about elections, wars and natural disasters.

Since May, websites hosting AI-created false articles have increased by more than 1,000 percent, ballooning from 49 sites to more than 600, according to [NewsGuard](#), an organization that tracks misinformation.

Most Read Technology >



- 1 Elon Musk's business empire is built on \$38 billion in government funding
- 2 See Spot live longer: Antiaging pill for dogs clears key FDA hurdle
- 3 American workers are skeptical AI will help them on the job
- 4 Online education company Chegg sues Google, says AI is killing business
- 5 Move fast, break things, rebuild: Elon Musk's strategy for U.S. government

Fake News

A growing issue

Since May 2023, the number of websites hosting fake-news has increased dramatically. Part of the issue lies in AI allowing for the creation of misleading or false articles with ease, using tools like chatbots and large language models.

The rapid proliferation of these sites, combined with the general lack of media literacy within the public, makes it difficult for readers to distinguish between what is real and fake

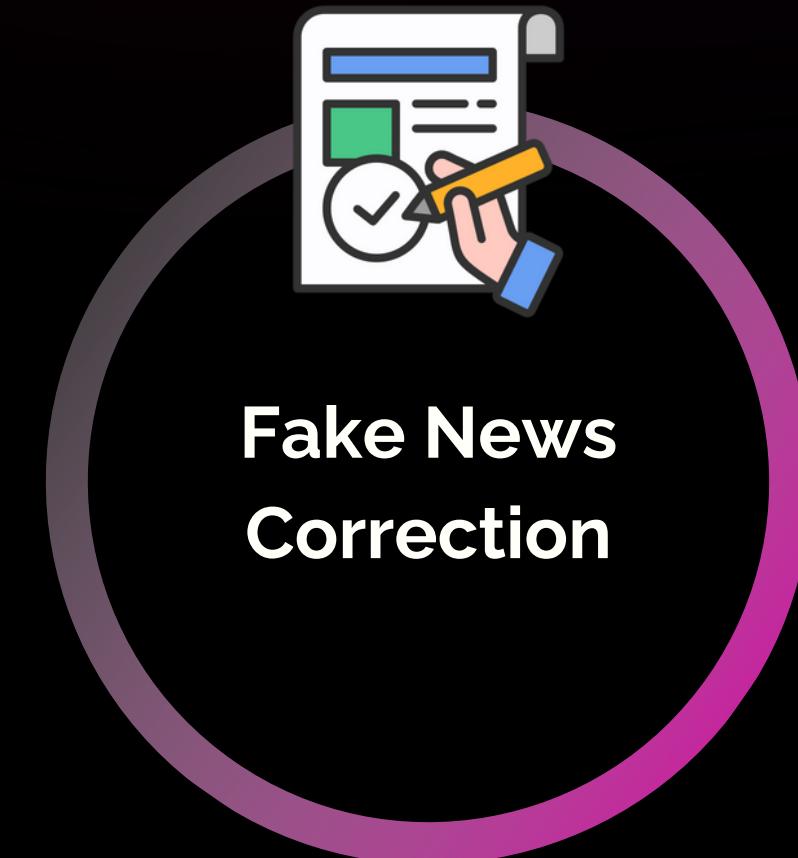
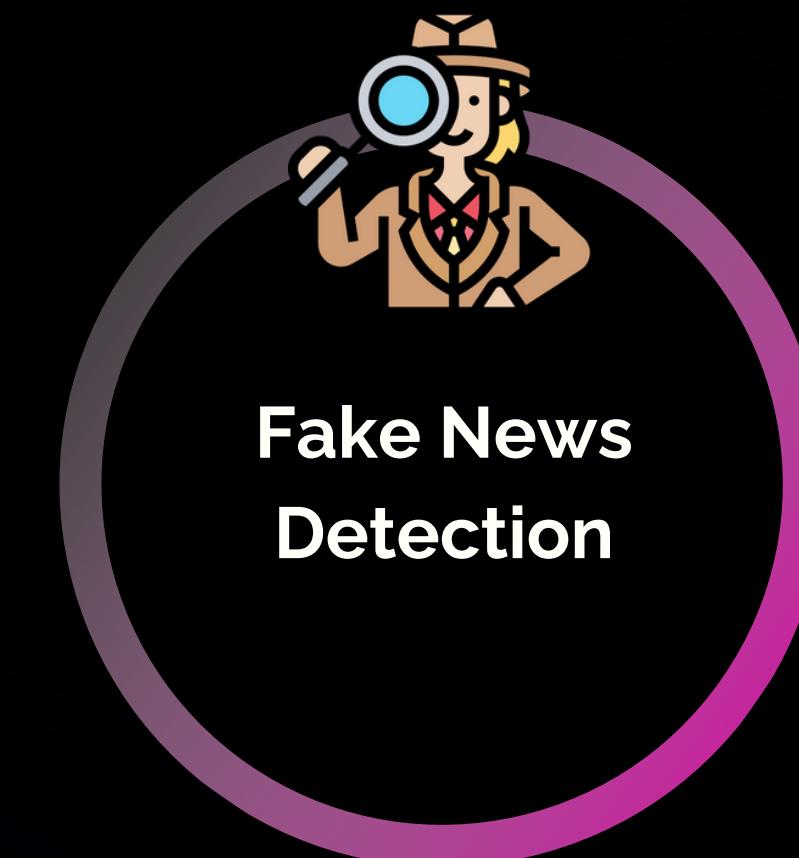
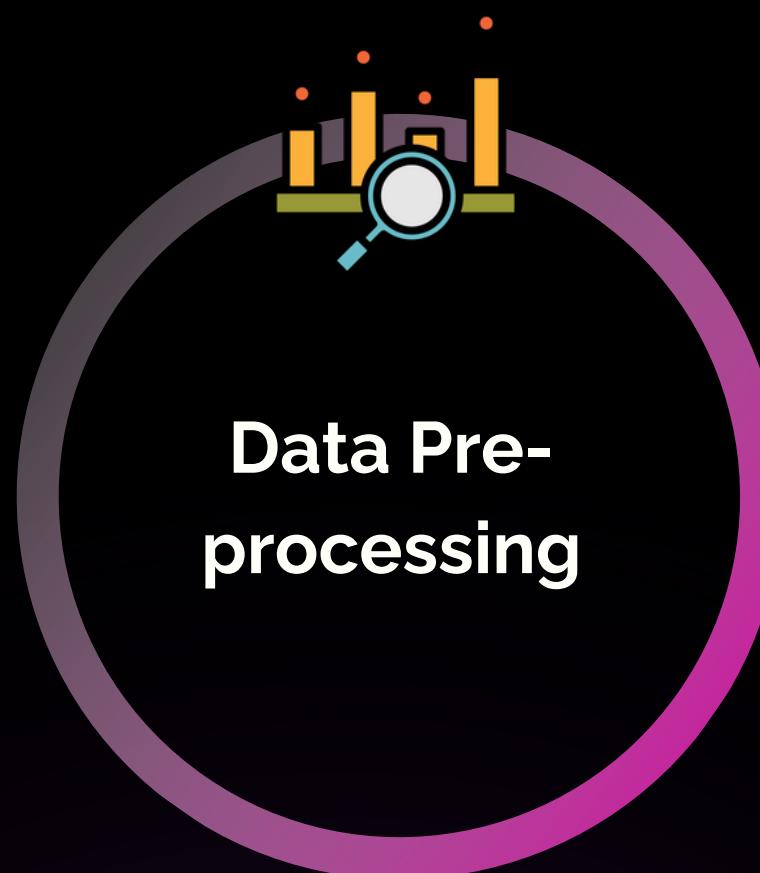
Even if readers were able to tell them apart, there is currently no solution out there to provide them with the truth - there's a real chance they'll click on yet another fake site



Team Empty String

Introducing EMBR!

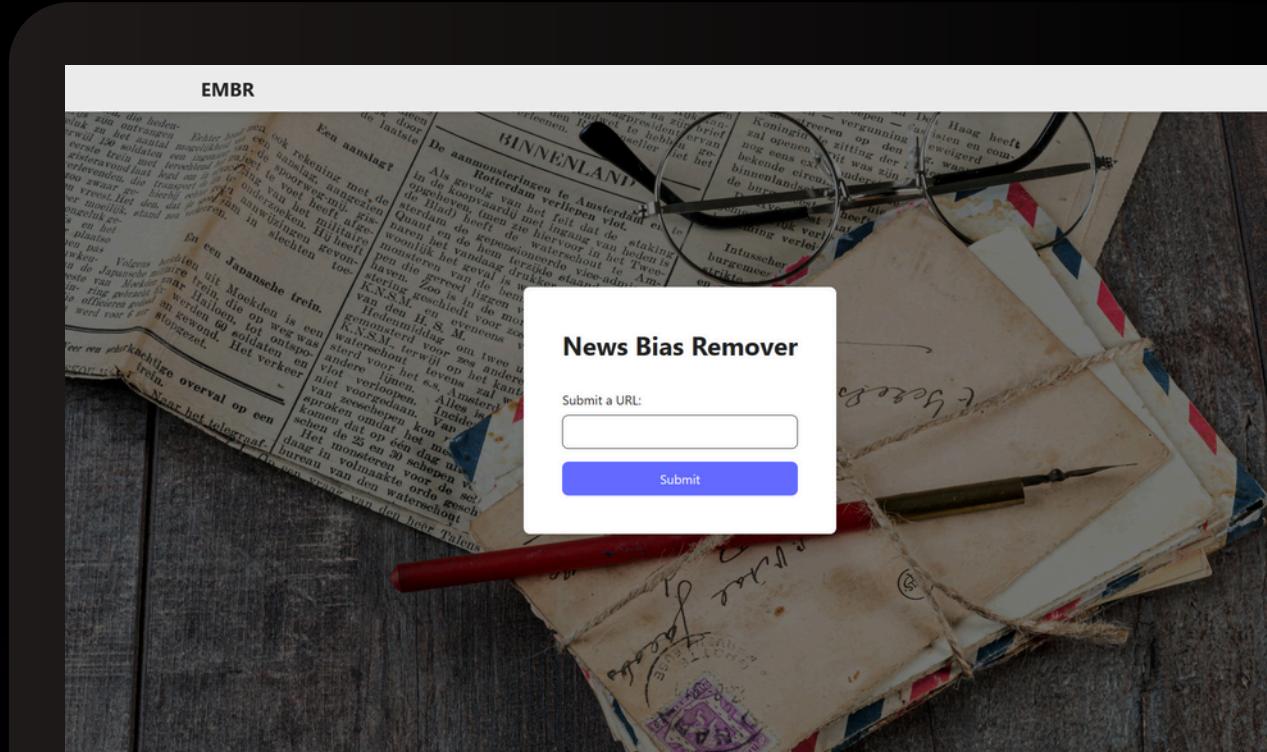
Enhanced Multimodal Bias Retrieval / Eliminator



EMBR does more than just detect fake news, it corrects it for the user too!



User Experience



Single Step Process
Simply Key in the URL of the article
you want to process - that's all!

Image

EMBR extracts all images from a webpage, then selects the relevant ones

Text

EMBR also parses the contents of the article

Provenance

EMBR is also able to identify the provenance of an article (e.g. Date Published, Source, etc)

Output

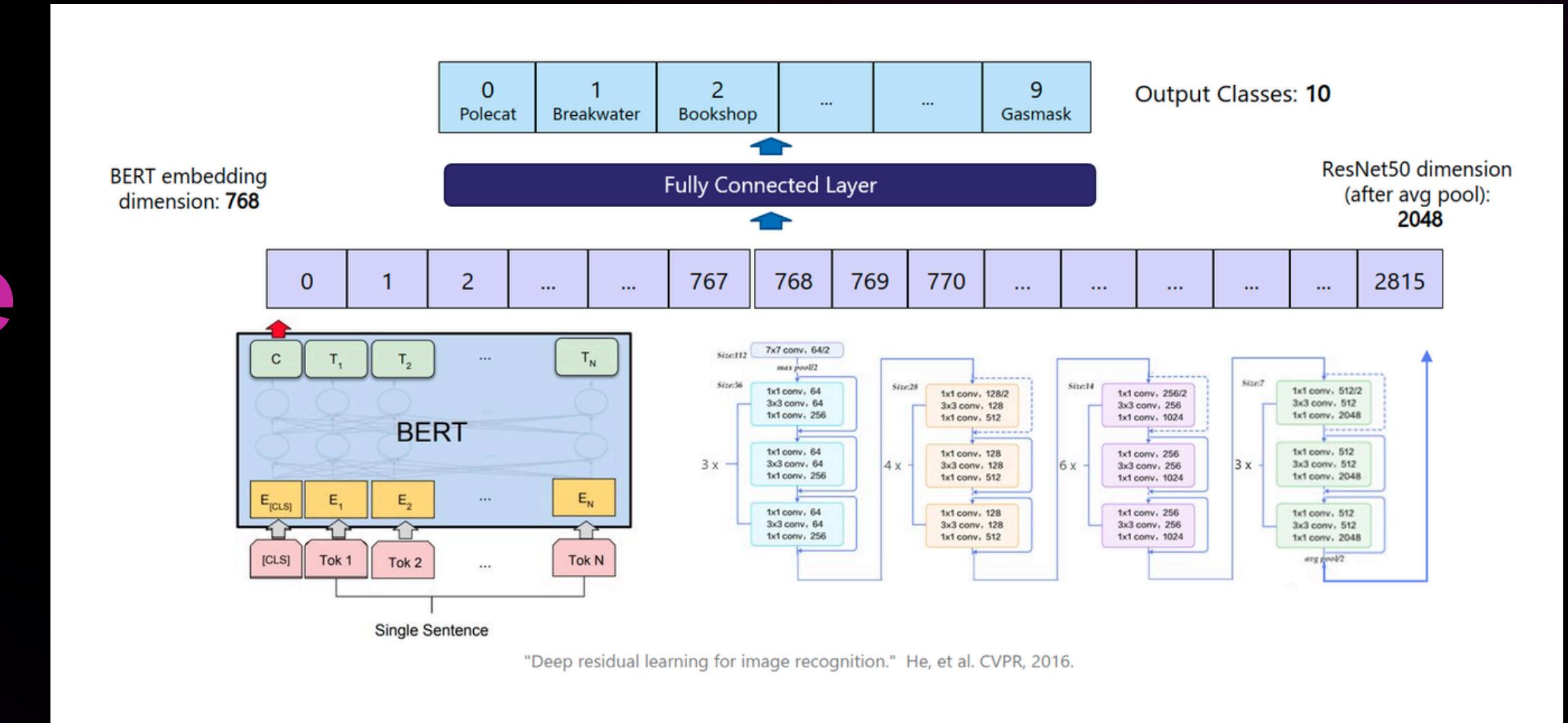
All of this gets formatted into JSON, which is passed to the Bias Model



Model Architecture

Inspired by He et al CVPR, 2016

Early Fusion





Team Empty String

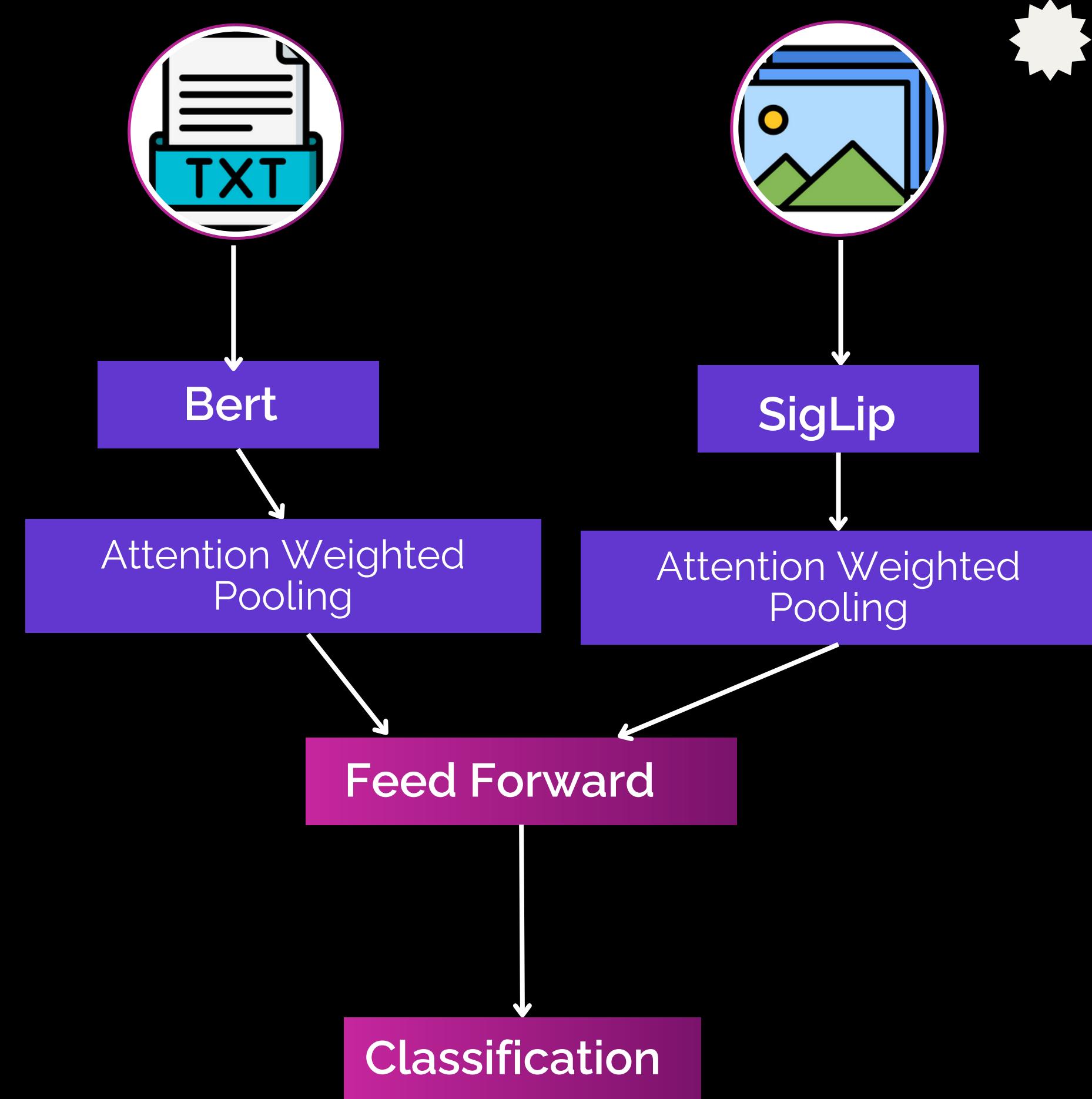
Machine Learning

After pre-processing, text is fed into Bert, while images are fed into SigLip

The model first determines whether the news is fake, then flags out sections of the article that are (with associated reasons)

This model was trained on labelled image and photo pairs

<https://github.com/trxs-7/EMBR>





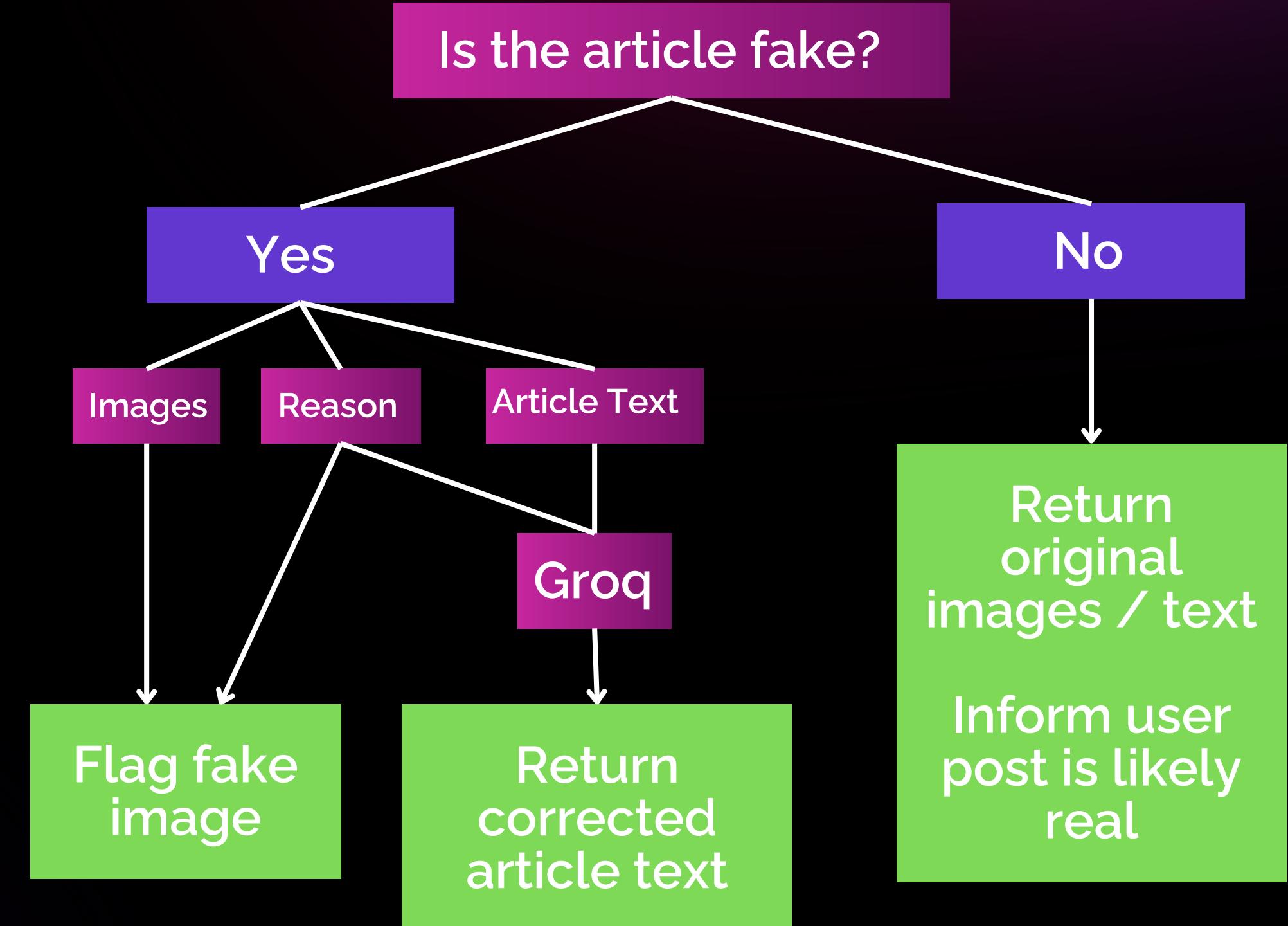
Team Empty String

Fake News Correction

If the article is flagged as fake, the pre-processing section will select relevant portions of the article (together with the context) to feed into a LLM for correction

The corrected information is then displayed to the user

For images flagged as false, clarification will also be given to explain the decision





What makes EMBR Unique?

■ Competitive Performance

Unlike models like Bert, our model can take in both images and text without any performance loss. In fact, it outperforms single medium models!

Our only real competitors are other LLMs, which are cost-prohibitive to implement for this purpose and are typically closed sourced + non explainable unlike our solution.

■ CI/CD Pipeline

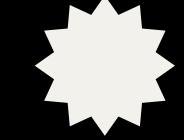
One issue with fake news databases is that they have to be continuously updated to stay relevant.

This project has also been built with Kubeflow Pipelines in mind, allowing for continuous integration and development from collaborators.

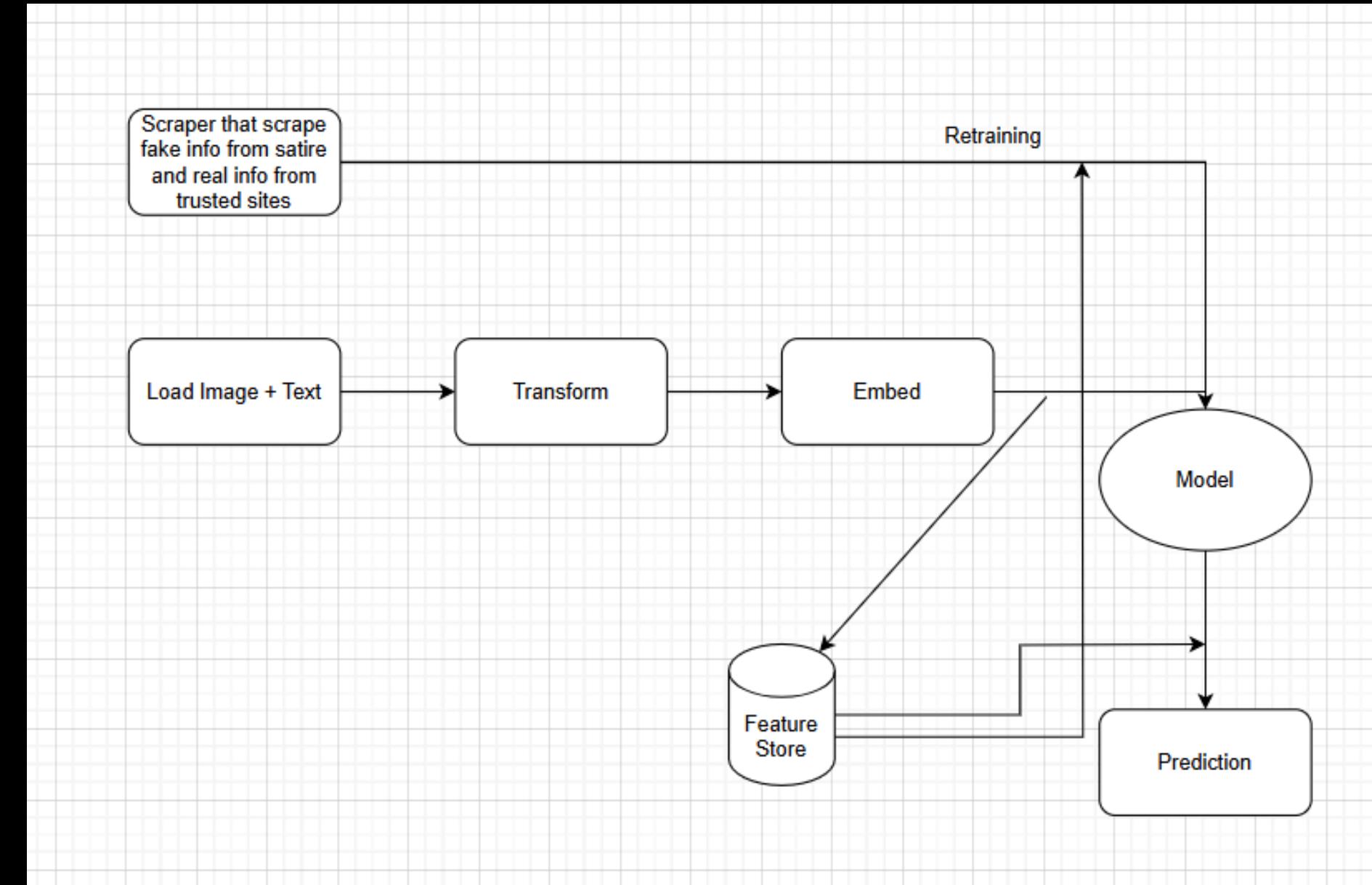
This is in addition to the improved security, reliability and performance offered by Kubeflow in the first place

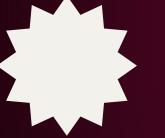
But wait, there's more...





Proposed CI/CD Pipeline





Limitations & Improvements

Data Recency

How accurate our models are at correcting the fake news depends on how when they were trained

=

Increased overhead from frequent updates

Expanded Scope

Features like audio recognition are easy to add, together with integrated gradients to examine the model's predictions

This will allow it to become an all-in-one tool for combating misinformation



