



News Headline Generator

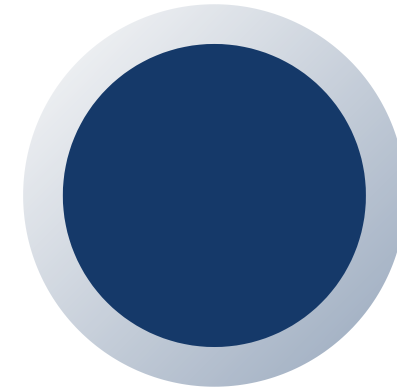
Arsh Jindal(202103021)

Aditya Tanna(202103023)

Vedant Pandya(202101063)

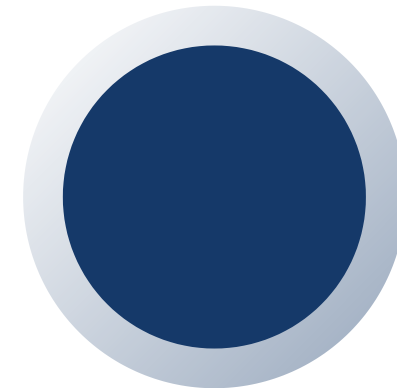
Approaches

We have approached the problem using three different unique ways



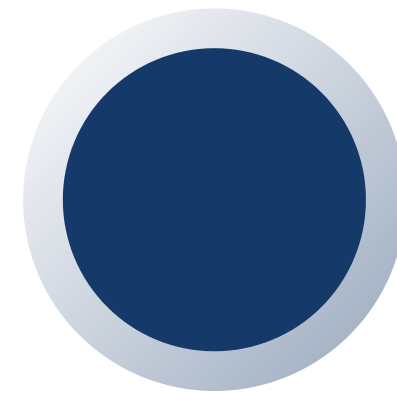
Transformer

Used a pre-trained T5 Transformer from huggingface library



LDA

Condensed article sentences into topics and selected the dominant ones to get representative

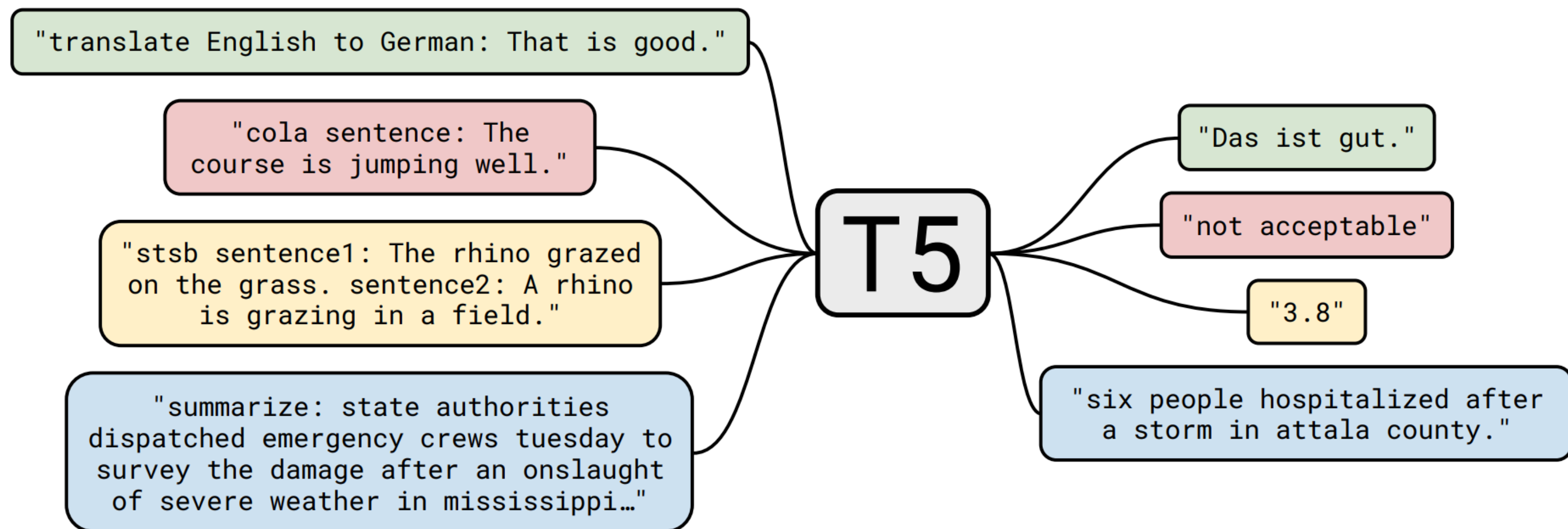


LSTM

Trained on the articles as input and the highlights as target

T5 - Transformer

- This approach generates summaries of text documents using the T5 transformer model and evaluates their effectiveness through comparison with summaries produced by another summarization technique (Sumy's LSA summarizer).
- The evaluation metric used is the ROUGE score, which assesses the overlap of content between the generated summaries and reference summaries



LDA - Approach

This is a unique approach.

- We take an article and with the sentences in it as documents, perform LDA on it to get the topics.
- Using the topic weight matrix, we calculate the dominant topics out of which some (2 in our case) are selected. For each topic, we can find multiple sentences belonging to it.
- For generating a headline, we get 1 sentence per topic and concatenate them.

LSTM - Approach

- LSTM is trained on the articles as input and the highlights as target and the headline generation would be done using beam search.
- Using an LSTM trained on articles with highlights as targets, complemented by headline generation using beam search, offers a robust method for generating succinct and relevant headlines.
- LSTMs are perfect in capturing long-term dependencies within text, essential for understanding complex article contexts and distilling this into headlines

Accuracy Results - T5

```
Generated Headline: the young actor says he has no plans to fritter his cash away on fast cars, drink and celebrity parties. he will be able to gamble in a casino, buy a drink in a pub or see the horror film "Hos
Original Highlight: Harry Potter star Daniel Radcliffe gets £20M fortune as he turns 18 Monday .
Young actor says he has no plans to fritter his cash away .
Radcliffe's earnings from first five Potter films have been held in trust fund .
ROUGE Scores: {'rouge1': Score(precision=0.25862068965517243, recall=0.38461538461538464, fmeasure=0.30927835051546393), 'rouge2': Score(precision=0.19298245614035087, recall=0.2894736842105263, fmeasure=0.231578
----
Generated Headline: inmates with most severe mental illnesses are incarcerated until they're ready to appear in court. judge says arrests often result from confrontations with police. prisoners are wearing sleeve
Original Highlight: Mentally ill inmates in Miami are housed on the "forgotten floor"
Judge Steven Leifman says most are there as a result of "avoidable felonies"
While CNN tours facility, patient shouts: "I am the son of the president"
Leifman says the system is unjust and he's fighting for change .
ROUGE Scores: {'rouge1': Score(precision=0.30303030303030304, recall=0.20408163265306123, fmeasure=0.24390243902439027), 'rouge2': Score(precision=0.03125, recall=0.020833333333333332, fmeasure=0.025), 'rougeL':
----
Generated Headline: survivor: "i probably had a 30-, 35-foot free fall. And there's cars in the water, there's cars on fire. the whole bridge is down" a doctor jumped into his car and rushed to the scene in 15 mi
Original Highlight: NEW: "I thought I was going to die," driver says .
Man says pickup truck was folded in half; he just has cut on face .
Driver: "I probably had a 30-, 35-foot free fall"
Minnesota bridge collapsed during rush hour Wednesday .
ROUGE Scores: {'rouge1': Score(precision=0.3, recall=0.36585365853658536, fmeasure=0.3296703296703297), 'rouge2': Score(precision=0.16326530612244897, recall=0.2, fmeasure=0.17977528089887637), 'rougeL': Score(pr
----
Generated Headline: the polyps were removed and sent to the national Naval Medical Center in Bethesda, Maryland. results are expected in two to three days. vice president Dick Cheney assumed presidential power at
Original Highlight: Five small polyps found during procedure; "none worrisome," spokesman says .
President reclaims powers transferred to vice president .
Bush undergoes routine colonoscopy at Camp David .
ROUGE Scores: {'rouge1': Score(precision=0.17142857142857143, recall=0.25, fmeasure=0.2033898305084746), 'rouge2': Score(precision=0.029411764705882353, recall=0.043478260869565216, fmeasure=0.035087719298245605)
----
Generated Headline: new: "your admitted conduct was not only illegal, but also cruel and reprehensible," nfl commissioner says. new: "your admitted conduct was not only illegal, but also cruel and reprehensible,"
Original Highlight: NEW: NFL chief, Atlanta Falcons owner critical of Michael Vick's conduct .
NFL suspends Falcons quarterback indefinitely without pay .
Vick admits funding dogfighting operation but says he did not gamble .
Vick due in federal court Monday; future in NFL remains uncertain .
ROUGE Scores: {'rouge1': Score(precision=0.15384615384615385, recall=0.24390243902439024, fmeasure=0.18867924528301888), 'rouge2': Score(precision=0.0, recall=0.0, fmeasure=0.0), 'rougeL': Score(precision=0.10769
----
```

The ROUGE scores presented demonstrate a modest overlap with the original highlights, indicating that although the model captures key information at times, it often omits relevant details or includes extraneous content. We can fine tune this model further to enhance the ROUGE scores.

Accuracy Results - LDA

```
▷ eval(sent_list, test['highlights'][0])
[9] ✓ 0.0s Python

... Generated headline:
    "What's objectionable is the attempts to undermine international justice, not Palestine's decision to join a treaty to which over 100 countries have agreed."

Highlight from data:
    Membership gives the ICC jurisdiction over alleged crimes committed in Palestinian territories since last June .
    Israel and the United States opposed the move, which could open the door to war crimes investigations against Israelis .

... {'rouge1': Score(precision=0.4117647058823529, recall=0.24561403508771928, fmeasure=0.3076923076923077)}
```

The ROUGE scores, with a precision of 41.17%, suggest that the model can correctly identify relevant terms and concepts.

Justification for the models used

- **Robust Framework:** We leverage the T5 Transformer, renowned for its versatility and state-of-the-art performance in natural language understanding and generation tasks.
- **Capturing Dependencies:** Using LSTM, we are able to capture long term dependencies which are useful when generating headline from all the important information in the article.
- **Topic Modeling:** Using LDA. we were able to do topic modeling and extract the important topics in each article and get the related sentences as headlines capturing what was in the article.

Creativity in our Model

- **Hybrid Model Utilization: Combining various aspects:** T5 Transformers, LDA topic modelling, and LSTMs—to have different outlooks to the model such as contextual understanding from Transformers, thematic focus from LDA, and sequence learning from LSTMs.
- **Topic-Driven Summarization:** Our unique application of LDA to identify dominant topics within articles and generate representative sentences provides a novel method to distill complex information into meaningful headlines.
- **Custom Context Calculation:** We created a custom simple context layer in our LSTM model that calculates attention-like weights to capture context information between the article and the highlights.