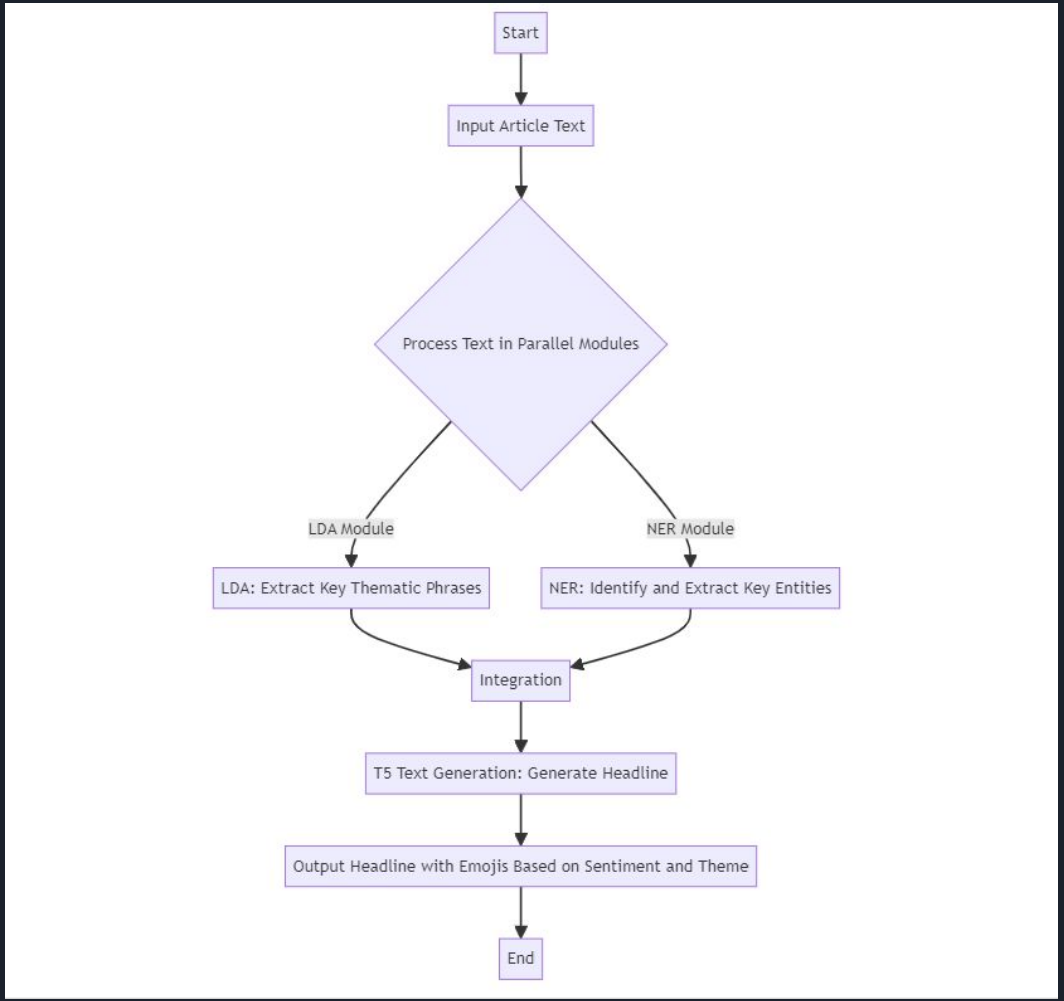
"Headline Generation Integrating LDA, NER, and T5"

"Exploring Creative Approaches for Accurate Headline Generation"

MODEL DESCRIPTION

This project uses advanced NLP technologies to automate and improve headline generation for news articles by integrating LDA, NER, and the T5 /MICHAUmodel. It extracts key topics and entities from articles, generates engaging and accurate headlines, and enhances readability by incorporating emojis through sentiment analysis .

.



WHY LDA ?

LDA (Latent Dirichlet Allocation)

Purpose: LDA is used to analyze the bodies of text to discover recurring patterns of words, which are grouped as topics.

Process: It treats each document as a mixture of topics and each topic as a mixture of words. This allows LDA to learn topic representations of documents and word representations of topics.

Contribution to Headlines: By extracting key phrases that represent the main themes of the articles, LDA ensures that the most relevant and significant topics are highlighted in the headlines, making them informative and contextually relevant.

WHY NER ?

Purpose: The T5 model is designed to handle a variety of text-based tasks within a single framework by converting all text-based language problems into a text-to-text format.

Process: T5 is pre-trained on a multi-task mixture of unsupervised and supervised tasks, optimized by minimizing the maximum likelihood across all tasks. This training enables it to understand and generate human-like text based on the input provided.

Contribution to Headlines: When provided with enriched inputs (article text combined with key phrases and entities), T5 generates coherent, contextually appropriate headlines. The ability to train on a customized dataset means it can adapt to the specific style and tone preferred by a news org

WHY MICHAU MODEL ?

The Michau/t5-base-en-generate-headline model is selected for its specialization in headline generation. Built on the T5 architecture, it is optimized to produce concise and impactful headlines, crucial for engaging readers quickly. Supported by robust research and community updates, the Michau model not only improves the efficiency of generating high-quality headlines but also adapts to the evolving needs of media and journalism, making it an ideal choice for this application.

Purpose: The T5 model is designed to handle a variety of text-based tasks within a single framework by converting all text-based language problems into a text-to-text format.

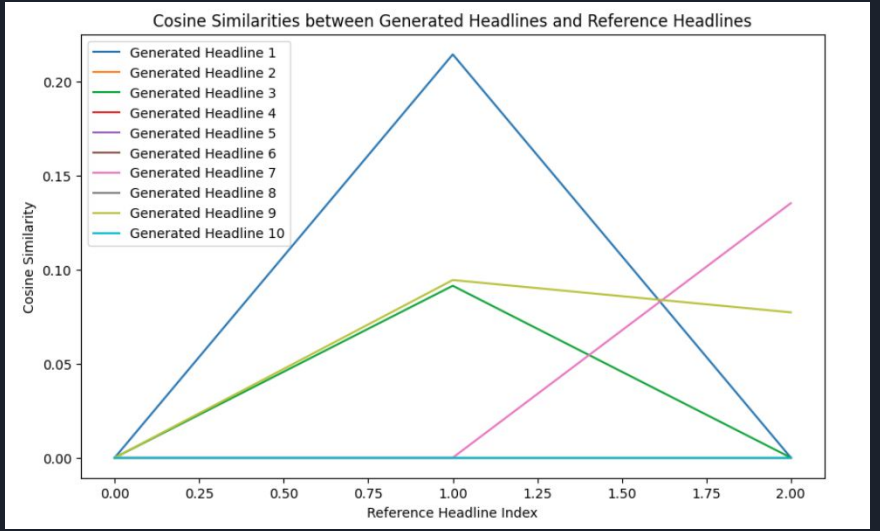
Process: T5 is pre-trained on a multi-task mixture of unsupervised and supervised tasks, optimized by minimizing the maximum likelihood across all tasks. This training enables it to understand and generate human-like text based on the input provided.

Contribution to Headlines: When provided with enriched inputs (article text combined with key phrases and entities), T5 generates coherent, contextually appropriate headlines. The ability to train on a customized dataset means it can adapt to the specific style and tone preferred by a news org

RESULTS

```
1 print(article)
2 print("Generated Headline:", headline)
```

```
WASHINGTON (CNN) -- As he awaits a crucial progress report on Iraq, President Bush will try to
Generated Headline: Bush to Invoke History to Defend America in Iraq 😊 🐰 [WASHINGTON]
```




```
1 generated_headlines
```

```
['Harry Potter star Daniel Radcliffe says he won\'t be a "Kid Star" 😢 [LONDON]',
 'Miami\'s "Forgotten Floor" 😢 [Editor]',
 'The Minneapolis Bridge Collapse 😢 [MINNEAPOLIS]',
 "President Bush's Colonoscopy - Doctors Remove Small Polyps From Bush's Colon 😢 [WASHINGTON]",
 'Atlanta Falcons quarterback Michael Vick is suspended without pay. Vick admits to participating in a dogfighting ring. 😢 [CNN]',
 'Youssif Says Burn 😢 [BAGHDAD]',
 "Prostitution in Iraq - A Woman's Choice 😢 [BAGHDAD]",
 "Colombian Defense Ministry Says 'El Negro Acacio' Was Killed 😢 [BOGOTA]",
 'White House Press Secretary Tony Snow to Step Down on September 14 😢 [WASHINGTON]',
 'New York City Police and FBI Investigating New Launcher Tube 😢 [CNN]']
```