

General Online Opinion Detector

G.O.O.D.

Members: William Coryell, Myung Lee, and
Ethan Cheung
Dr. Yilmaz
Computer Systems Lab
5/21/2025

-



The Solution

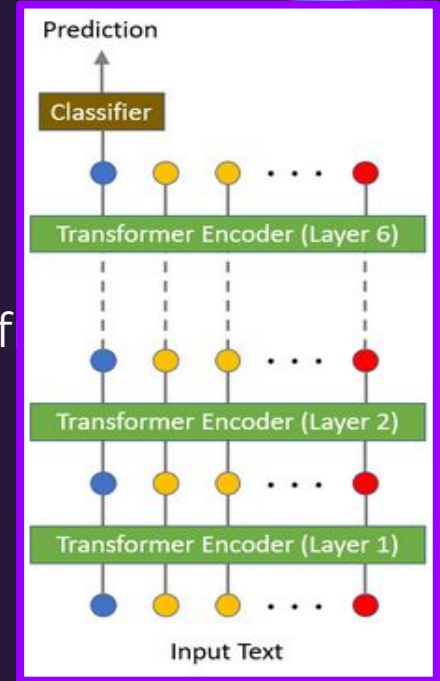
- Sentiment Analysis on multiple articles and internet posts
- Finds opinions based on topic
- Displays to user



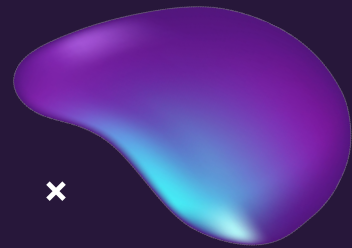
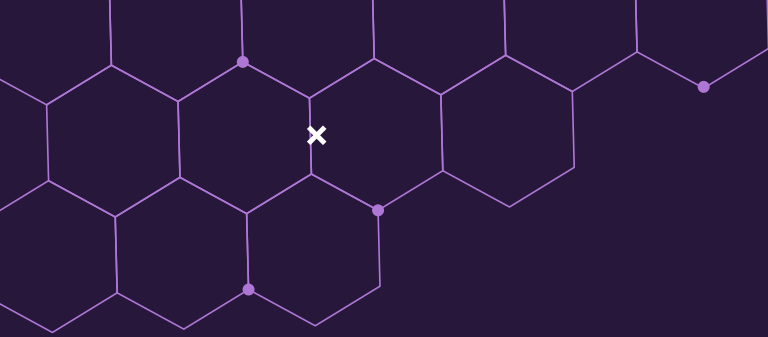
Background

Sentiment Analysis

- ML model that determines
 - Subjectivity
 - Polarity
 - Intensity
- BERT (Bidirectional Encoder Representation from Transformers)
 - Language model for NLP
- Web Scraping
 - Automated website data extraction



BERT Architecture



Other Solutions



Attention-based model using RNN [15]

×

- Used attention mechanisms to weigh word importance with RNNs to determine sentiment
- Limited because they do not use dropout rates to prevent overfitting.

×

+



Gradient-boosted support vector machine (GBSVM) [16]

×

- Utilized GBSVMs to analyze the sentiment of user reviews on popular social media platforms.
- Applications are constrained to datasets with only tf-idf(term frequency measurement) of uni-, bi-, and tri-gram as features.

×

+



Sentiment analysis with deep CNNs and sequential algorithm [17]

×

- Extracted aspects of words and speech vectors to train a deep CNN model
- Employed a sequential algorithm to obtain a sentiment annotation of a sentence.
- Limited performance with computing power.

×

+



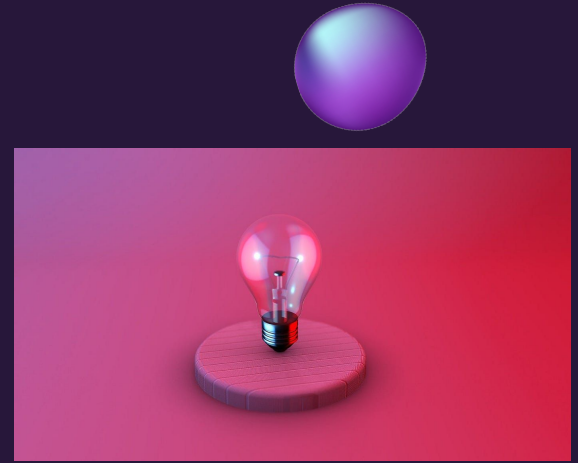
Why is ours better?

- Aspect-based sentiment analysis
- Train data for multiple forms of media
 - X(Twitter) and News Articles
- Compiles multiple sources of data
- User-friendly



Novelty

- Finds its own data points, rather than needing to be fed individual post/articles.
- The user friendly website
 - Has graph displaying sentiment using media posts and new outlets as data points
- Extracts data from websites, rather than needing input



Impact

x



x

- Echo chamber detection
- Polarized topics for political focus
 - What issues for politicians to target
- Determine amount of concern public holds for a subject
- Opinion popularity across internet

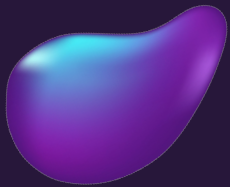
x

Methods

- Website takes a topic as a input
- Finds multiple articles relating to that topic
 - Web Scraping raw text
- Detects the opinions of the articles on the topic
 - Aspect-Based Sentiment Analysis, DeBERTa
- Displays each articles opinion on a graph

Methods

- Aspect-Based Sentiment Analysis
 - DeBERTa
 - PyTorch
- Web-Scraping
 - Google News and Article API
- Website
 - Flask
 - HTML/CSS/JS

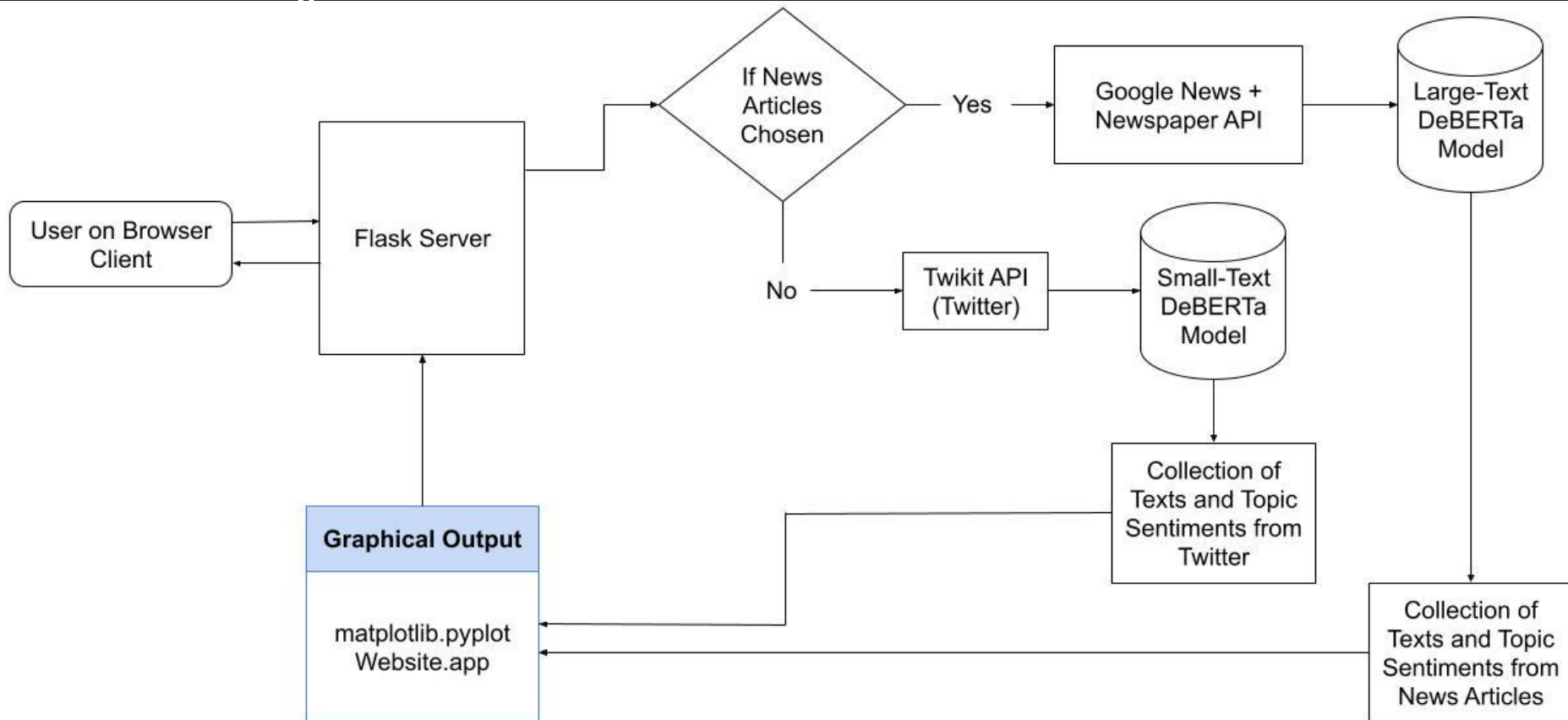


x



Methods: Workflow

x

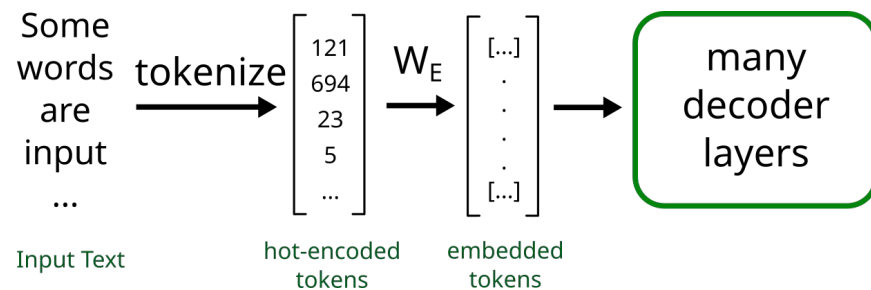




BERT Process

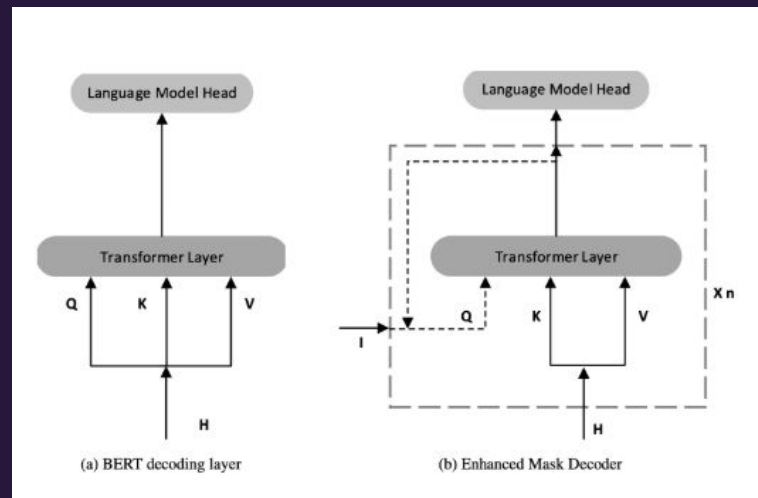


“Tokens”



Decoding-enhanced BERT with disentangled attention (DeBERTa)

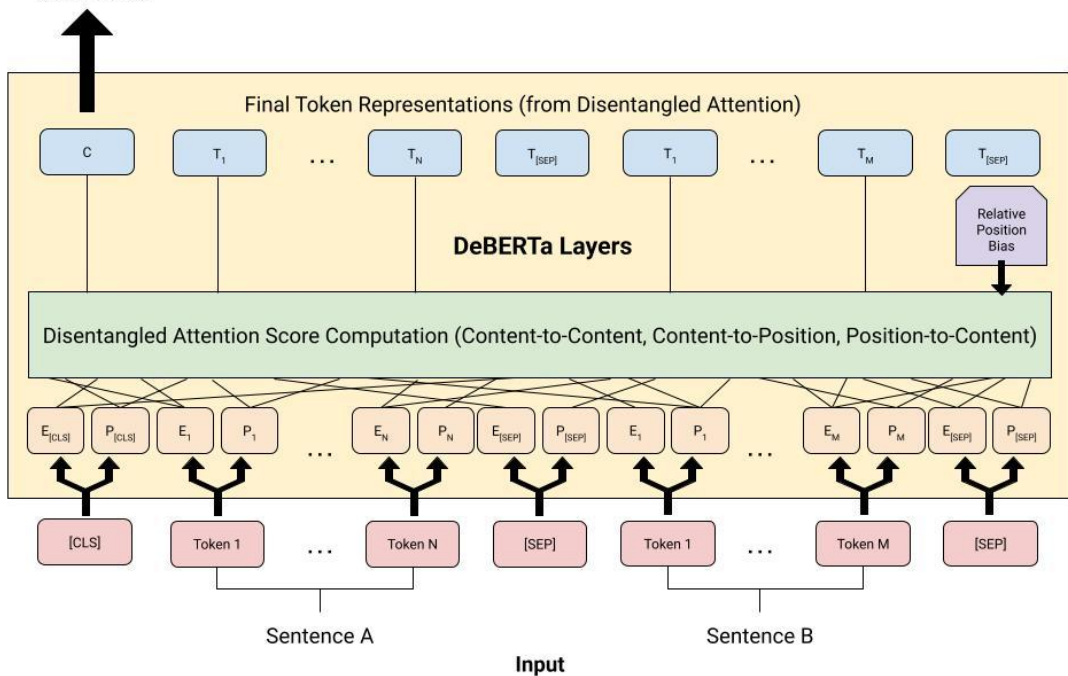
- Separates content and position into two vector encodings instead of combining them
- Enhanced Mask Detection with relative position in transformer layers instead of absolute
- BERT understands meaning of words using context, whereas DeBERTa also understands their relationships
- DeBERTa v3
 - Replaced token detection (RTD)
 - ELECTRA-Style pre-training
 - Gradient Disentangled Embedding Sharing



Model Architecture

DeBERTa v3

Class Label



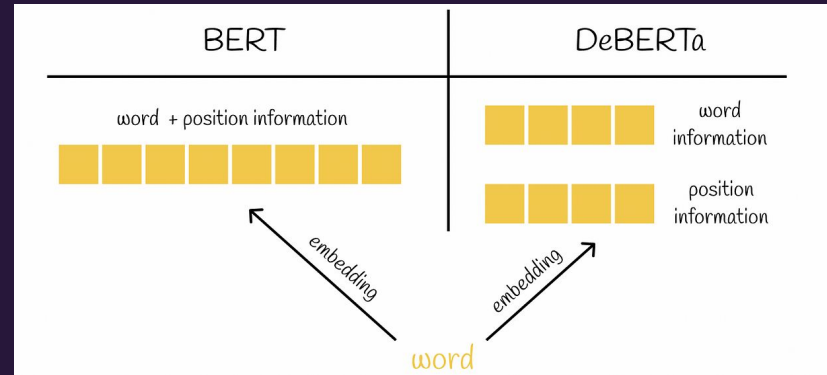


Methods - Model Training

- Downstream tasks in NLP
 - BERT- pre trained on Wikipedia and Google's BooksCorpus (3.3 billion words)
 - Combination of Transfer Learning and Self-supervised learning to fine tune pretrained models
 - Allows BERT to find sentiment of aspects without having to relearn word embeddings or contextual understanding
- DeBERTa model
 - 10 Epochs
 - **Short text dataset:** ~1900 entries of X(Twitter) and Yelp Reviews, using 80-20 train test split
 - **Long text dataset:** ~10000 entries of text from articles and online reviews
 - Aspects labeled Positive, Neutral, or Negative
 - Accuracy and F1 score calculated on test set after each epoch

Small Text Model Results

- Microsoft/DeBERTa v3 base model on Hugging Face trained with PyABSA library dataset
- ~85% accuracy and F1 score of 0.81
- Classifies sentiment of words within text using embeddings consistently and accurately
- Scaled to have similar impact as large text model



Small Text Model Results (cont.)

- Transformed model output
- Specify terms to classify through using [B-ASP] and [E-ASP] embeddings
- Things to improve:
 - Model initialization time
 - More variety in dataset
 - Other measurements, such as precision and recall
 - Further tuning hyperparameters
 - Include dropout rates

```
[2024-11-06 11:30:50] (2.4.1.post1) Example 0: But the <staff:Negative Confidence:0.9985> was so horrible to us .
[2024-11-06 11:30:50] (2.4.1.post1) The results of aspect term extraction have been saved in C:\Users\ethan\mygit\ASBA-DeBERTa\Aspect Term Extraction and Polarity Classification.FAST_LCF_ATEPC.result.json
[2024-11-06 11:30:50] (2.4.1.post1) Example 0: Not only was the <food:Positive Confidence:0.9992> outstanding , but the little ` <perks:Positive Confidence:0.9973> \ ' were great .
[2024-11-06 11:30:51] (2.4.1.post1) The results of aspect term extraction have been saved in C:\Users\ethan\mygit\ASBA-DeBERTa\Aspect Term Extraction and Polarity Classification.FAST_LCF_ATEPC.result.json
[2024-11-06 11:30:51] (2.4.1.post1) Example 0: It took half an hour to get our <check:Neutral Confidence:0.9946> , which was perfect since we could sit , have <drinks:Neutral Confidence:0.9987> and talk !
[2024-11-06 11:30:51] (2.4.1.post1) The results of aspect term extraction have been saved in C:\Users\ethan\mygit\ASBA-DeBERTa\Aspect Term Extraction and Polarity Classification.FAST_LCF_ATEPC.result.json
[2024-11-06 11:30:51] (2.4.1.post1) Example 0: It was pleasantly uncrowded , the <service:Positive Confidence:0.9989> was delightful , the <garden:Positive Confidence:0.9989> adorable , the <food:Positive Confidence:0.9944> - LRB - from <appetizers:Positive Confidence:0.9632> to <entrees:Positive Confidence:0.978> - RRB - was delectable .
[2024-11-06 11:30:52] (2.4.1.post1) The results of aspect term extraction have been saved in C:\Users\ethan\mygit\ASBA-DeBERTa\Aspect Term Extraction and Polarity Classification.FAST_LCF_ATEPC.result.json
[2024-11-06 11:30:52] (2.4.1.post1) Example 0: How pretentious and inappropriate for MJ Grill to claim that it provides <power lunch:Neutral Confidence:0.9989> and <dinners:Neutral Confidence:0.9991> !
```

Input Text: The service is excellent , the decor is nice and elegant.

Aspect: service
Sentiment: Positive 😊
Confidence: 98.96%

Aspect: decor
Sentiment: Positive 😊
Confidence: 98.95%

Input Text: This burger has by far the worst taste in the entire continent.

Aspect: taste
Sentiment: Negative 😞
Confidence: 97.07%

Input Text: What is the meaning of this?

Aspect: meaning
Sentiment: Neutral 😐
Confidence: 79.62%



Large Text Model Results

- microsoft/deberta-v3-**large** base model trained on Yelp reviews and a political sentiment analysis datasets.
- Slightly less accurate (~80% accuracy)
 - Tested on test dataset of ~1000 entries
 - Trade off for being faster for larger text



Website Design

[home](#)[github](#)[about](#)

General Online Opinion Detector

brought to you by GMI

a DeBERTa solution to polarization and opinion

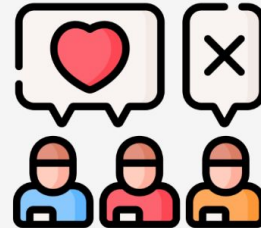
Find Varied Opinions on Topics Across the Internet

Enter Topic:

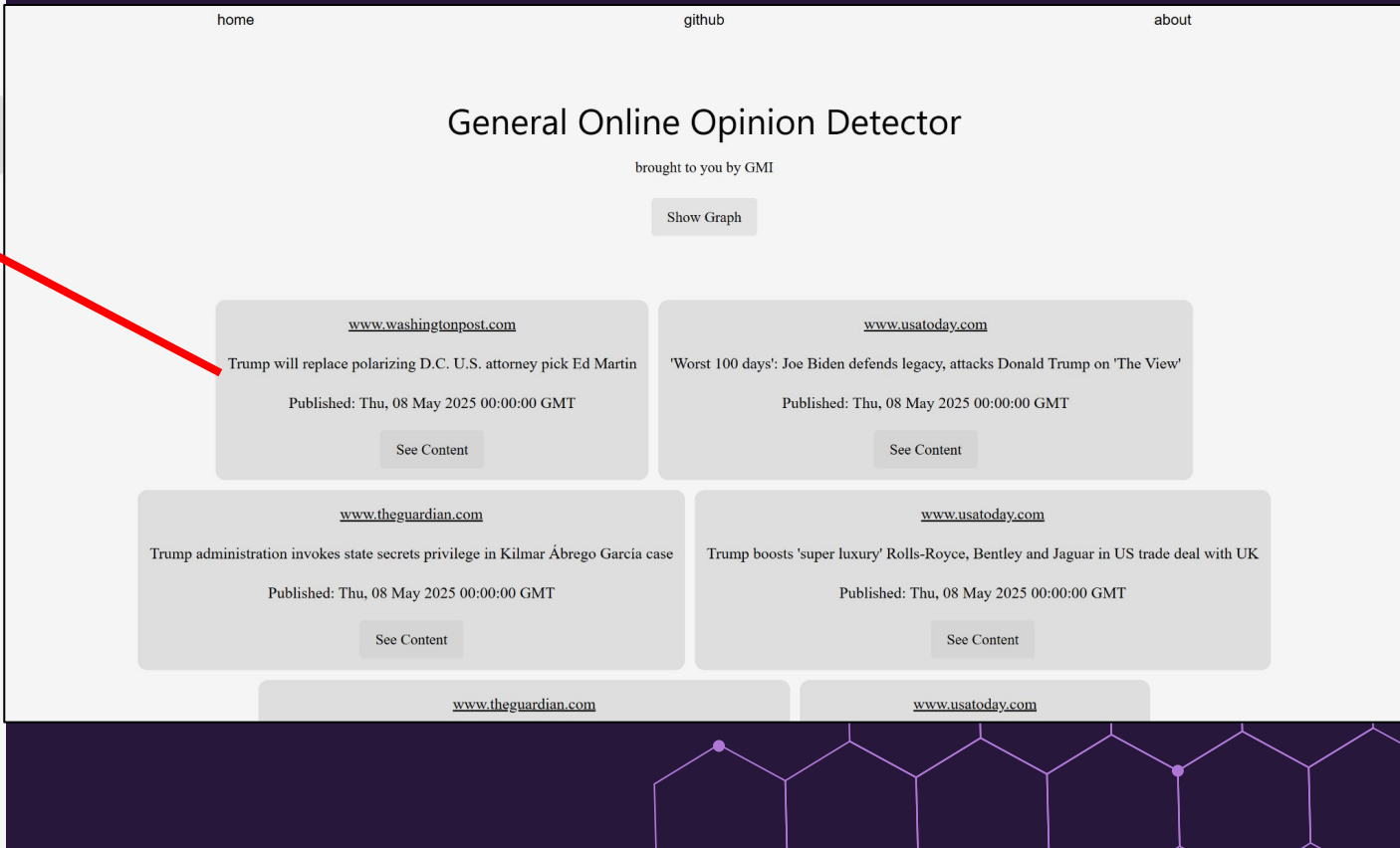
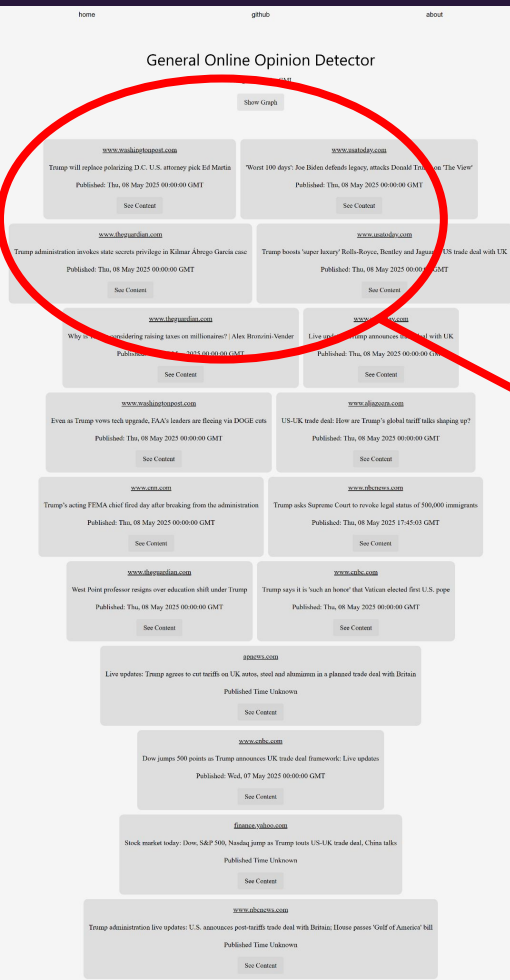
Select Source:

News/Long Text Model ▾

Submit



Website Design (cont.)



Website Design (cont.)

home

github

about

General Online Opinion Detector

brought to you by GMI

Content from: <https://apnews.com/live/donald-trump-news-updates-5-8-2025/>

Recent deadly plane crashes and technical failures that have put a spotlight on the outdated

U.S. air traffic control network are prompting the Trump administration to propose an overhaul. The plan calls for six new air traffic control centers, along with technology and communications upgrades at all of the nation's air traffic facilities over the next three years

said Transportation Secretary Sean Duffy. How much it will all cost wasn't immediately revealed. "Decades of neglect have left us with an outdated system that is showing its age,"

Duffy said in prepared remarks. "Building this new system is an economic and national security necessity, and the time to fix it is now." The Trump administration wants to add fiber, wireless and satellite technology at more than 4,600 locations, replace 600 radars and increase the number of airports with systems designed to reduce near misses on runways. ►

Read more about the Trump administration seeking to overhaul US air traffic control

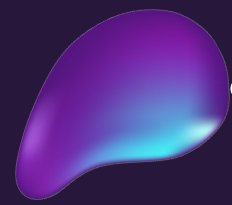
the sentiment is 0

Website Design (cont.)





Limitations

- 
- Time
 - Computing power
 - Only some articles
 - ABSA taking more data
 - Website takes too long to load
 - Model
 - Calculating cumulative overall sentiment
 - Hard to determine capability

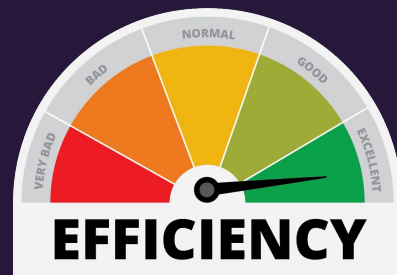
x

Future Work

×

- More varied social media sources for data
 - YouTube
 - TikTok
 - Instagram
- Reduce time limitations
 - Faster ABSA model
 - Quicker article and post fetching
 - Increase amount of articles used
 - Diversification of model

×

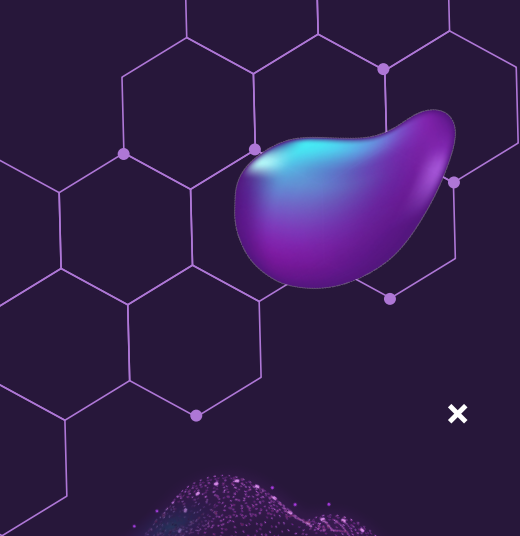


Conclusion

- Allows for readers to understand diverse perspectives
- Fosters learning through analyzing viewpoints
- Encourages the formation of unique opinions
- Presents the media in a readable format

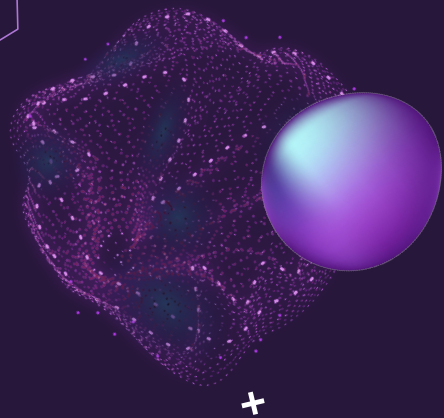


Q&A



x

THANKS!



+

x

