

KALEIDO: Top-K Retrieval of News with Interpretable Embedding and Diversified Ordering

Anthony Tung, Yiqun Sun, Yixuan Tang, Qiang Huang, Yuanyuan Shi, Yingchaojie Feng

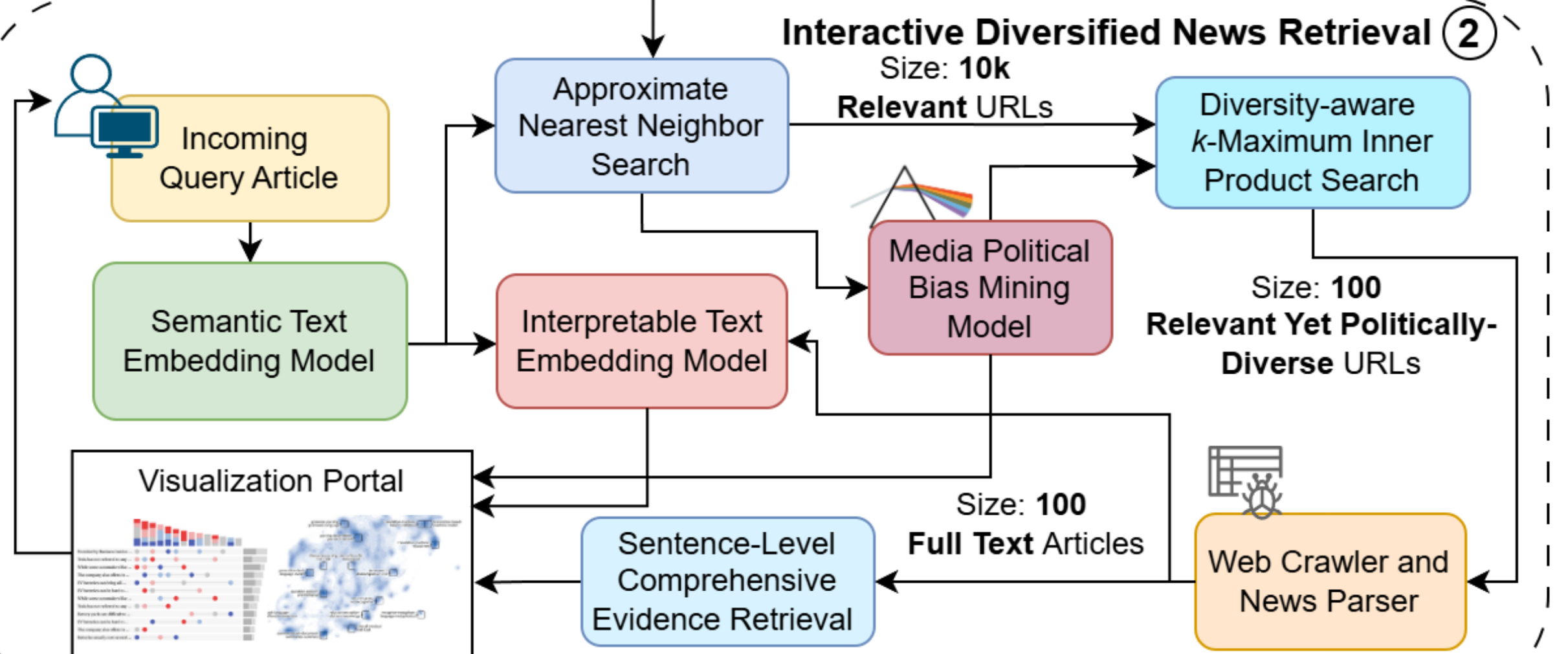
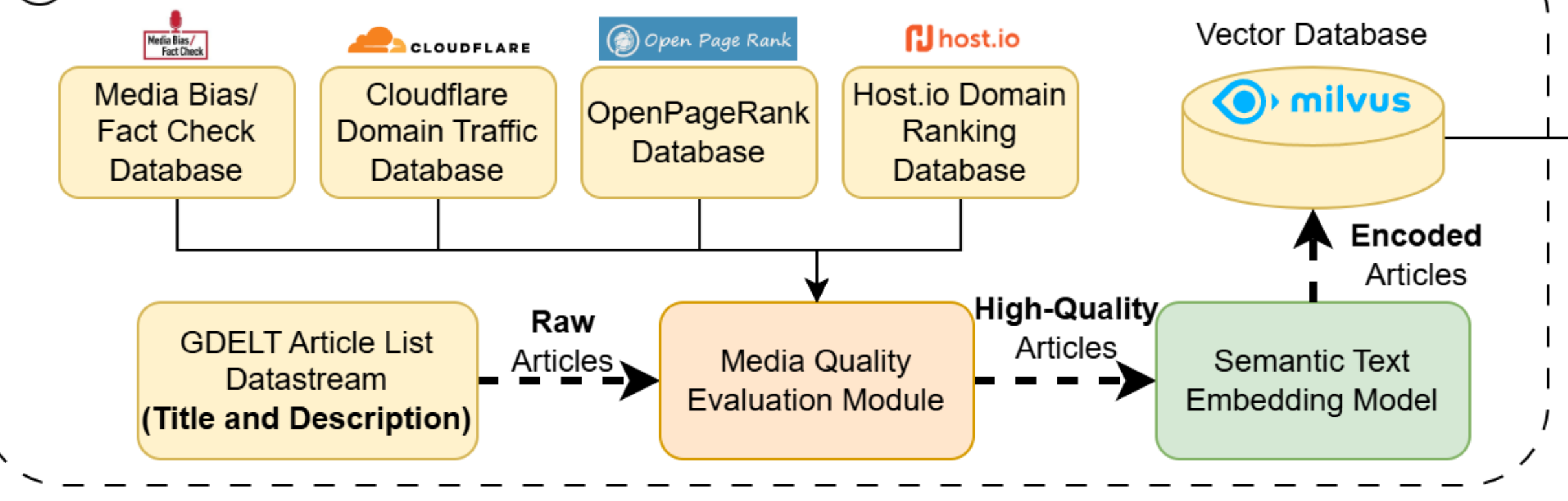
Centre for Trusted Internet and Community

National University of Singapore

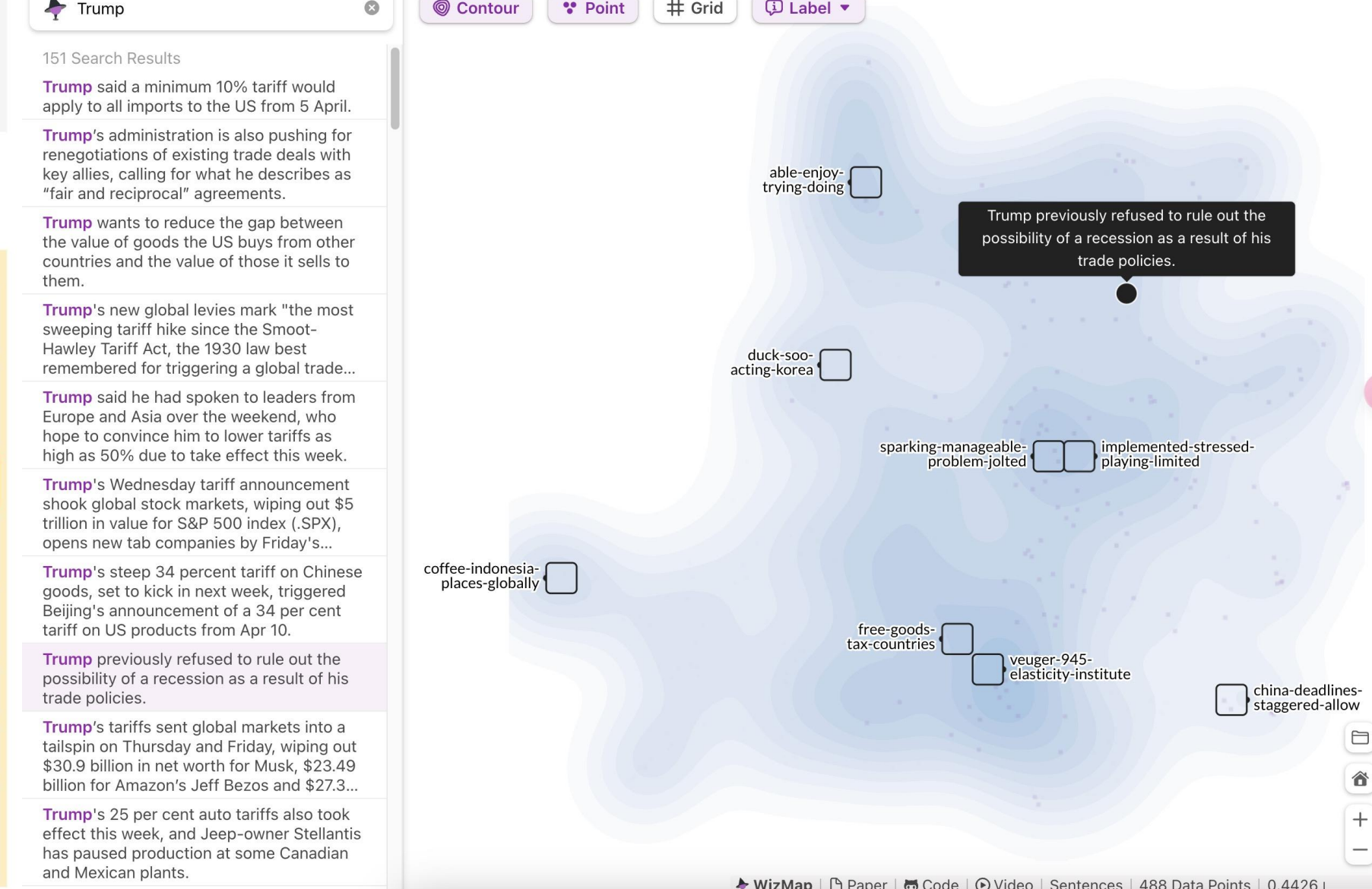
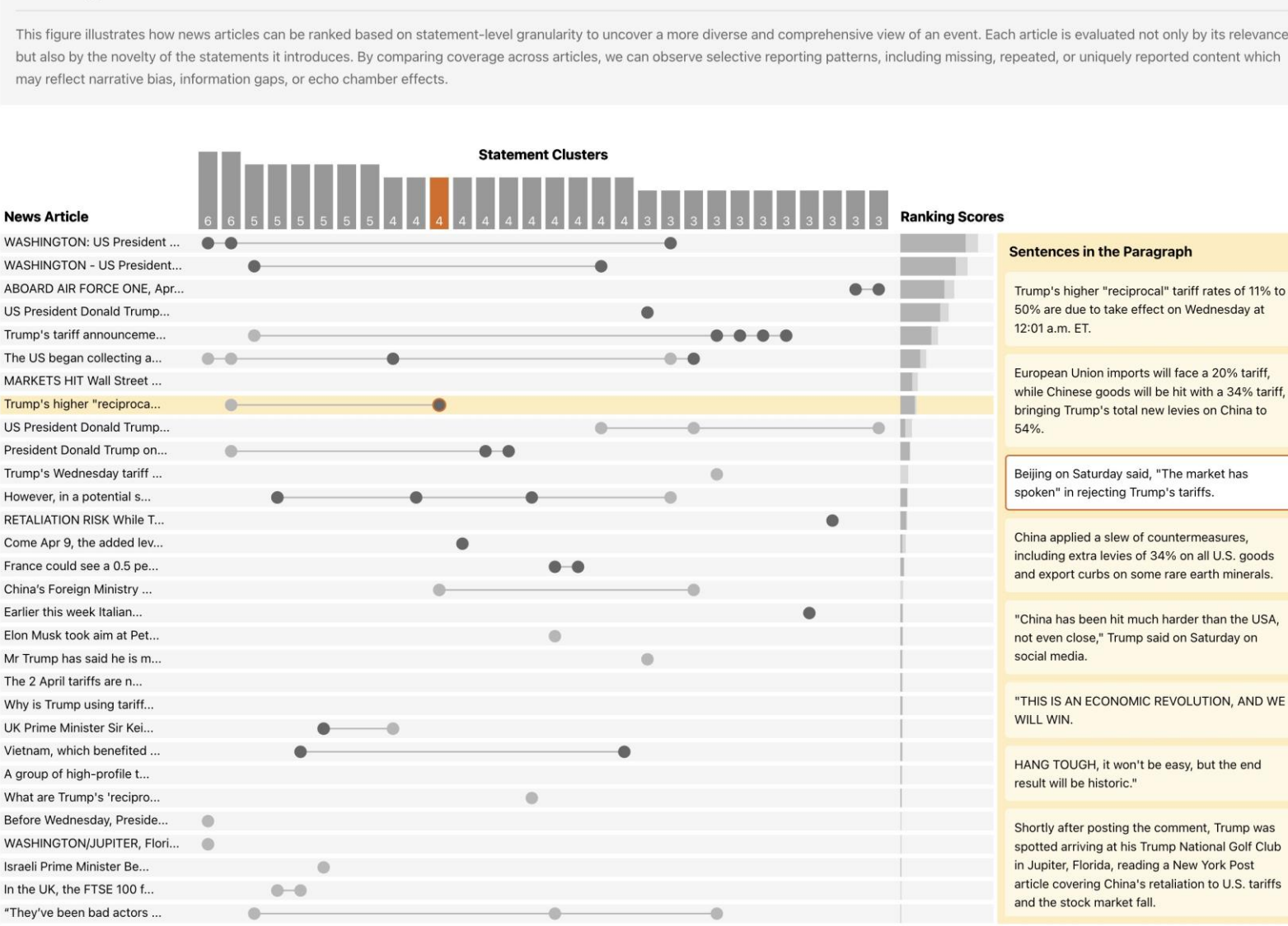


KALEIDO: Interpretable Diversified News Retrieval

1 News Articles Ingestion

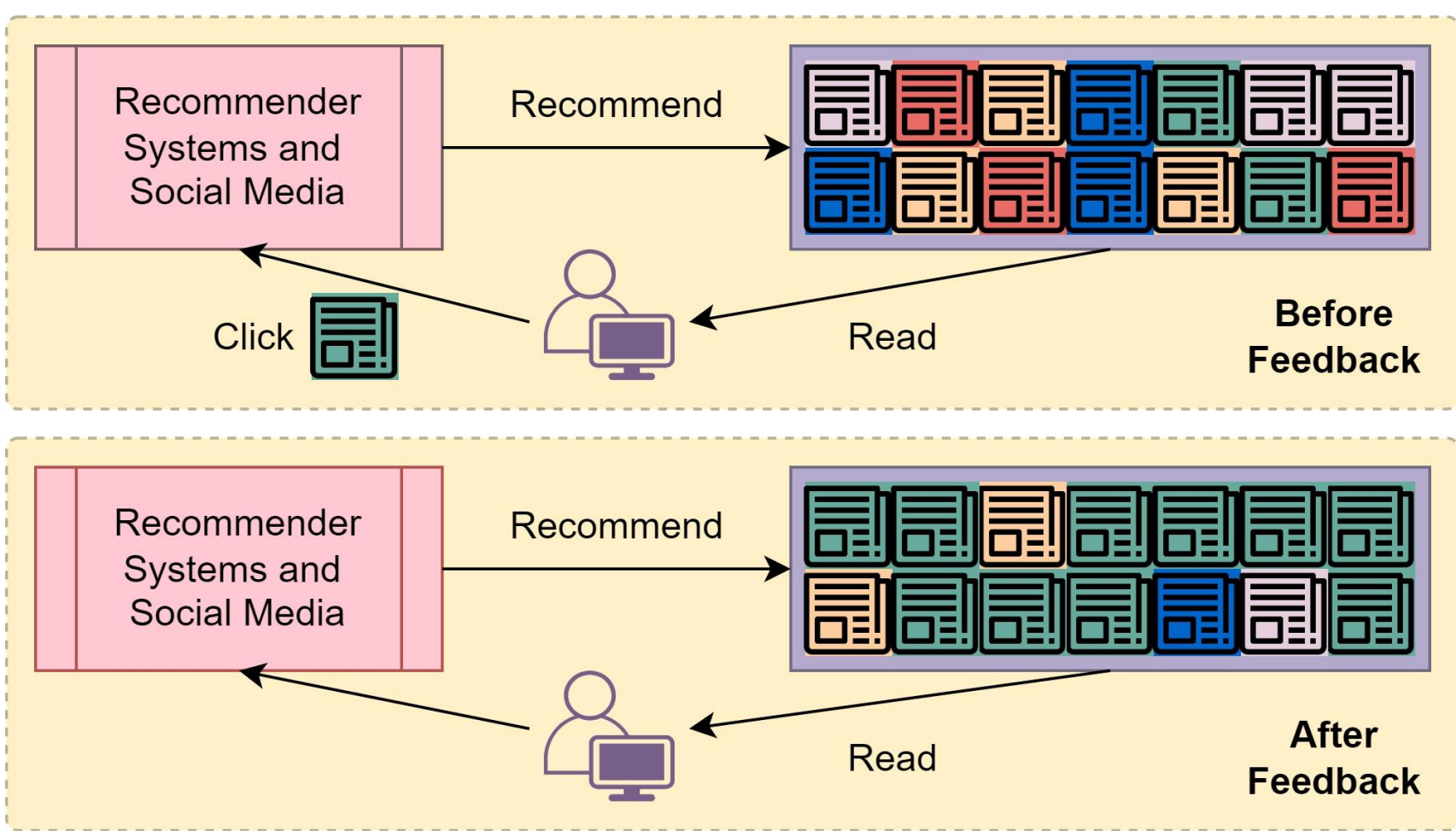


Coverage of Statements Across Retrieved Diverse Articles



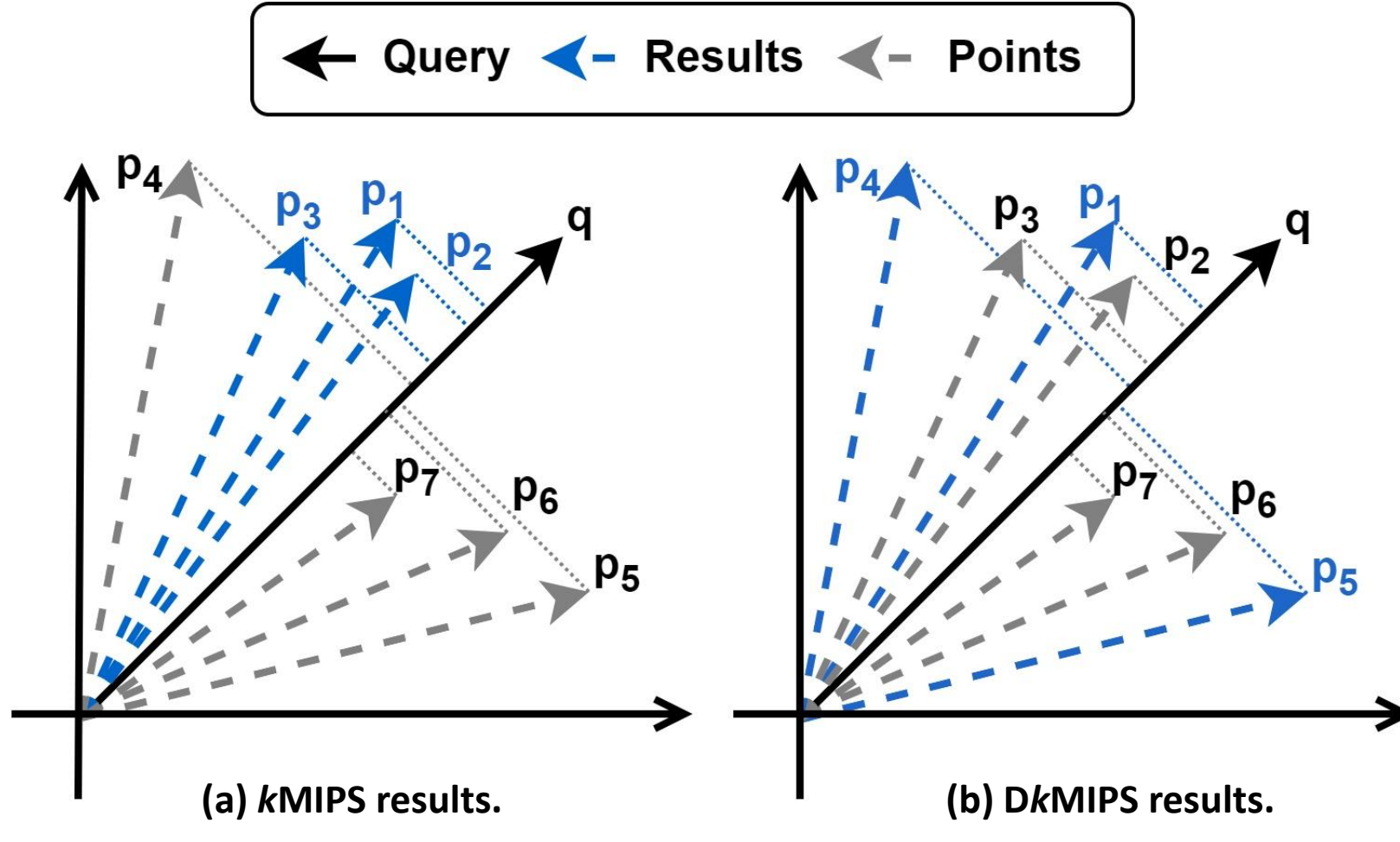
- A live, effective and user-friendly tool for comprehensive news consumption with interpretable diversified news retrieval.
- Sourced from real-word streaming GDELT data, updates every 15 mins, covers 152 languages and thousands of media websites.
- Politically-aware and interpretable diversification of the retrieval results.
- Provides interactive visualization tool for easy and comprehensive understanding of perspectives and events.
- Paragraph-Cluster view explains the reason of sentence-level diverse re-ranking.
- WIZMAP view visualizes the clusters of all retrieved statements.

Motivation: Echo Chamber Effect

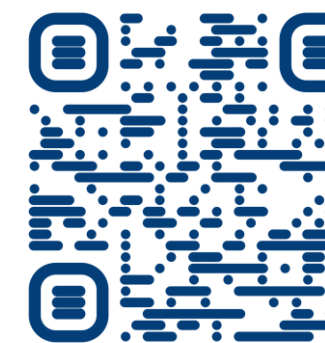


- Hidden topic/political preference is captured and reinforced in recommender systems and social media via the feedback loop.
- Results diversification helps news consumers to cross-validate and develop the ability to gain a comprehensive view.

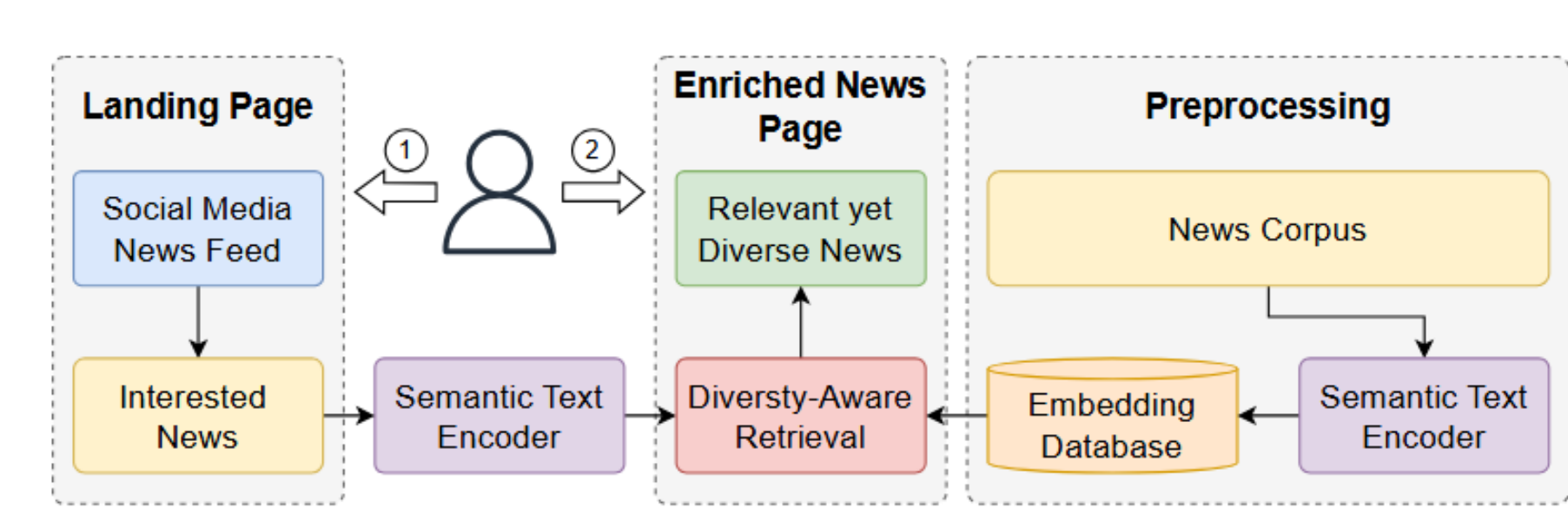
Diversity-Aware k-Maximum Inner Product Search (DkMIPS)



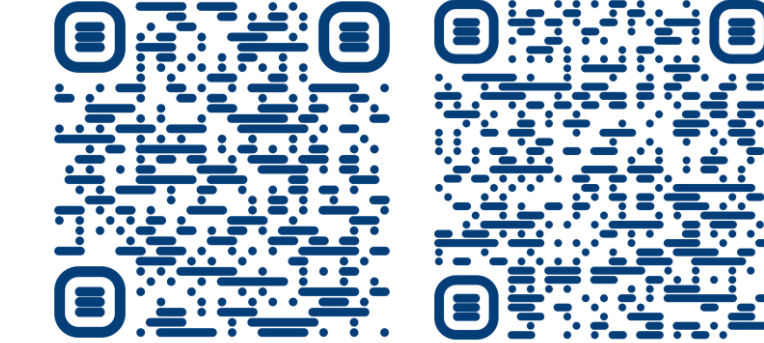
- DkMIPS provides a result set that balances relevance and diversity beyond vanilla vector similarity search, offering a off-the-shelf alternative to traditional MIPS e.g. FAISS.



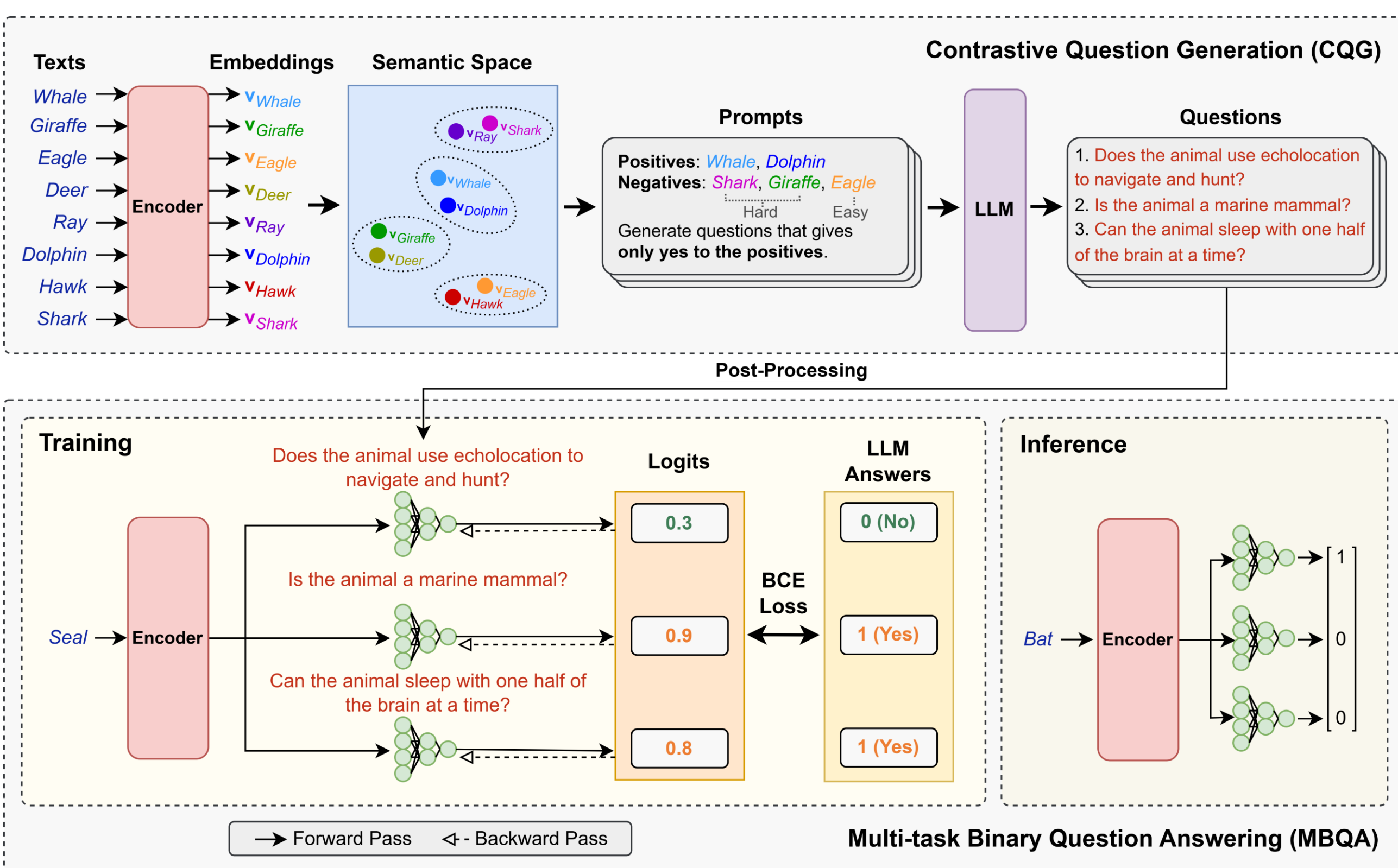
DiversiNews: Diversified News Reading



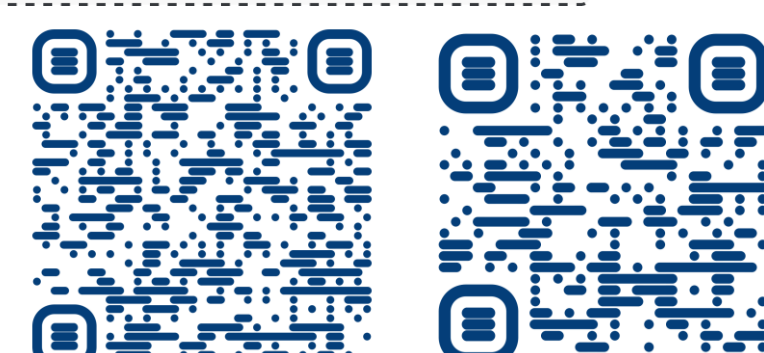
- DiversiNews employs semantic text embedding models and DkMIPS to provide relevant yet diverse news article in complement to the current reading.
- Proof-of-concept and pioneering work in enriching news reading with diversified and relevant articles.
- Still lacks interpretability and political awareness.



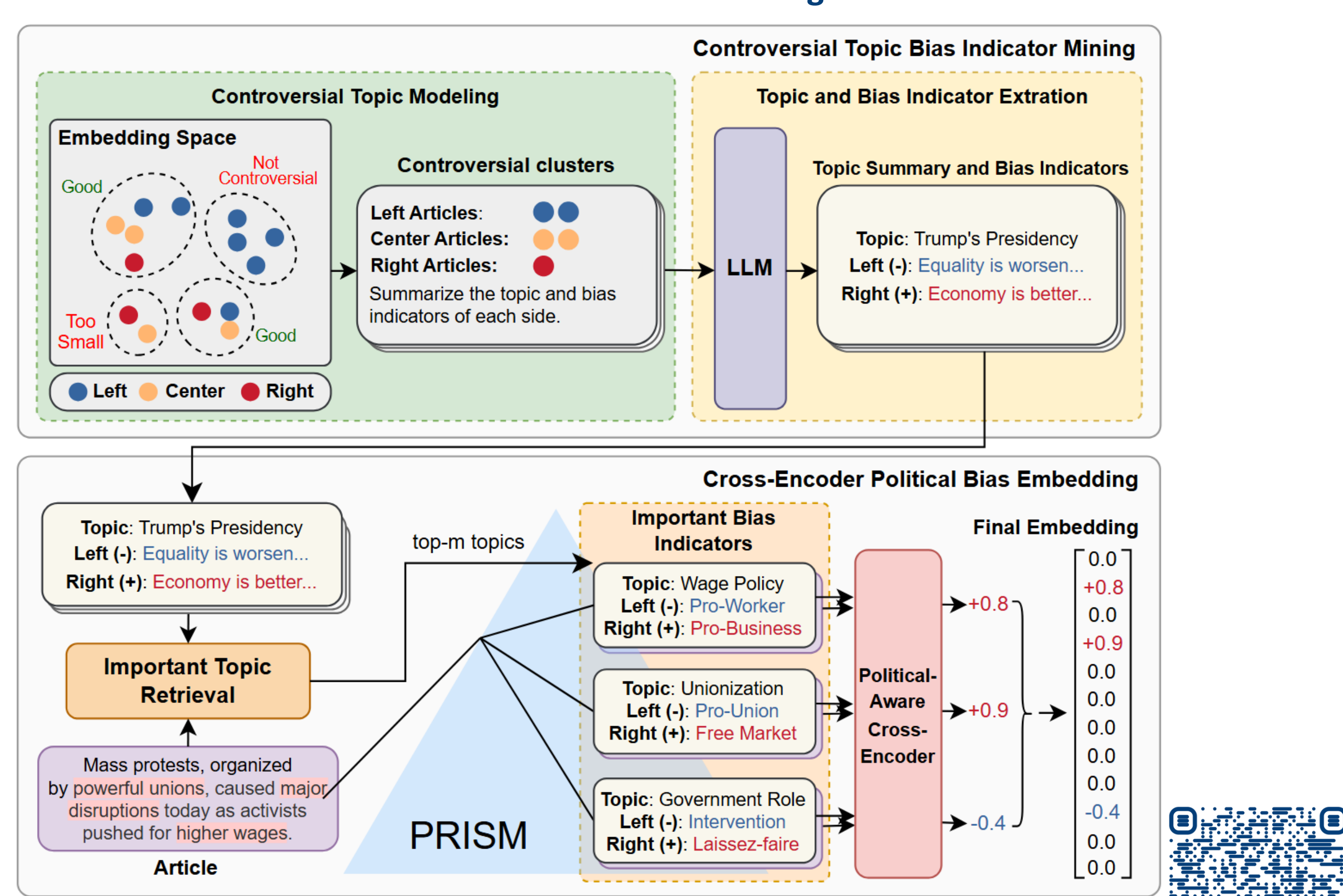
CQG-MBQA: Interpretable Text Embeddings



- Mining discriminative questions as interpretable embeddings with Contrastive Question Generation (CQG) and Efficiently answering them to form interpretable text embeddings with Multi-Task Binary Question Answering (MBQA).
- Effective in a variety of downstream tasks e.g. Retrieval, Classification.



PRISM: Political Bias Embeddings



- PRISM mines controversial topics and respective bias indicators to create fine-grained high-dimensional political bias representation of news articles, beyond traditional left-right scale.



Sentence-Level Comprehensive Diverse News Retrieval

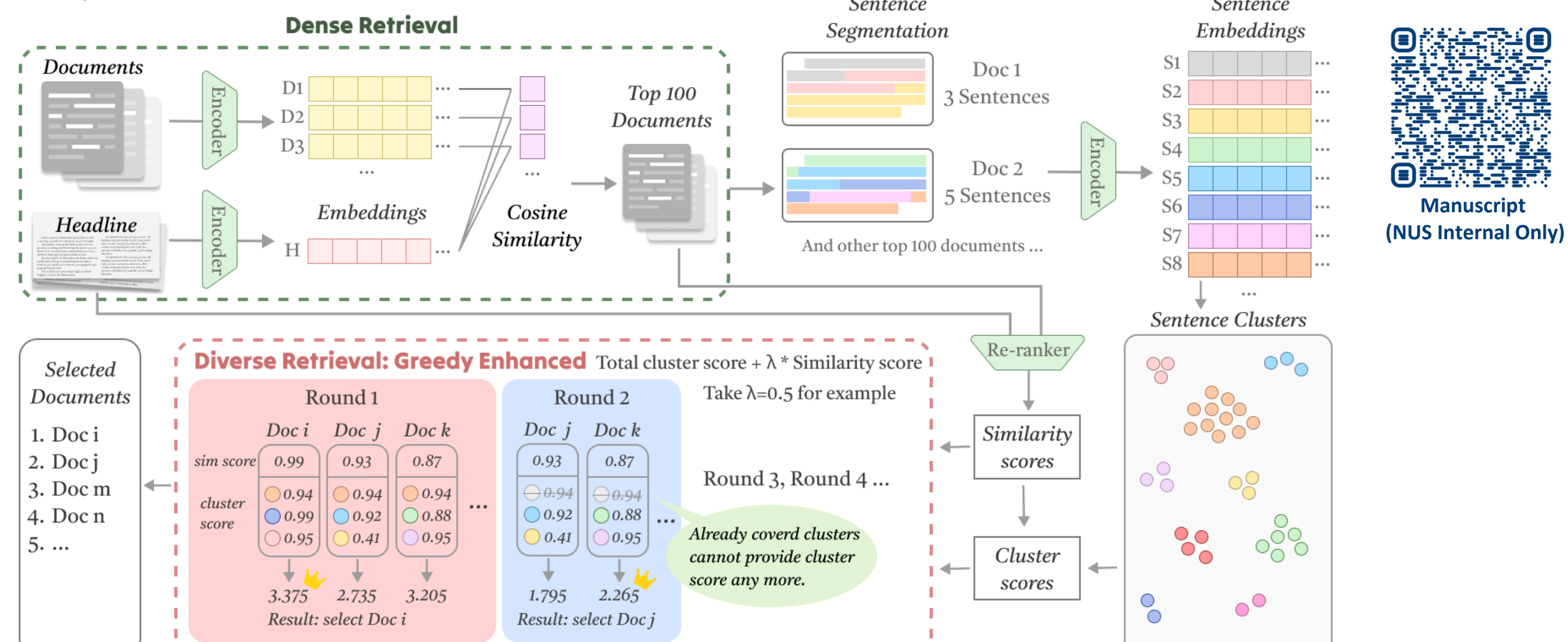
Search: Tesla's Model Y battery design faces criticism for "zero reparability".

<https://www.businessinsider.com/xxx...>
EV batteries can be hard to repair after accidents, so many often get scrapped. One expert said Tesl...

<https://www.msn.com/xxxx...>
Auto manufacturing expert Sandy Munro said the Tesla Model Y battery pack has "zero reparability..."

<https://www.yahoo.com/xxxxx...>
For many electric vehicles, there is no way to repair or assess even slightly damaged battery packs a...

Relevance Retrieval	After Diverse Re-Ranking
1 Experts said the Tesla Model Y battery pack has "zero reparability." ...	1 Experts said the Tesla Model Y battery pack has "zero reparability." ...
2 Its new structural battery pack has been described by experts as having "zero reparability." ... (Very similar to Rank 1)	2 Tesla has not referred to any problems with insurers writing off its vehicles. ...
3 EV batteries can be hard to repair after accidents, so many often get scrapped. ...	3 Tesla's decision to make battery packs part of the car's body has allowed it to cut production costs. ...
4 Tesla EV batteries are hard to repair after accidents, so many often get scrapped. ... (Very similar to Rank 3)	4 The company also offers its own insurance product in a dozen U.S. states to Tesla owners at lower rates. ...
5 Battery packs are difficult to repair and diagnose after even minor accidents. ... (Very similar to Rank 3)	5 For many electric vehicles, there is no way to repair or assess even slightly damaged battery packs after accidents. ...
Top-ranked paragraphs are duplicated / very similar.	More diversified, covering comprehensive aspects of facts and opinions.



- Treats each sentence as an atomic unit expressing facts or opinions about a news event. Enables fine-grained retrieval that captures multiple perspectives, overcoming the limitations of document-level similarity-based methods.
- Assigns relevance to the headline and diversity across sentence clusters, using a greedy algorithm to re-rank initial document-level retrieval results for interpretable and complementary selection.
- Achieves strong performance in surfacing hidden narratives and enhancing comprehensive understanding of events, ensuring retrieved articles reflect a broad spectrum of viewpoints rather than redundant reporting.