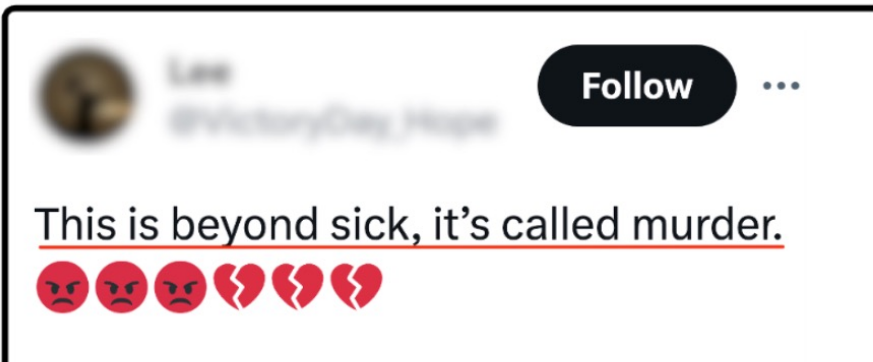


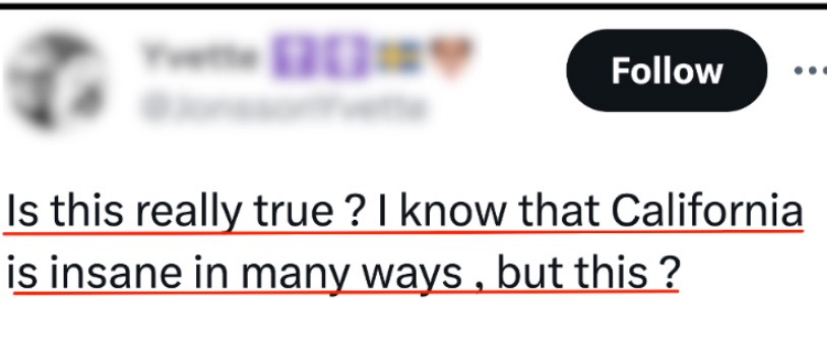
## Truthfulness Stance Detection

- Online information provides a valuable lens through which we can gauge people's perceptions and opinions, offering insights into societal trends, beliefs, and behaviors that shape human society.
- Truthfulness stance: given a factual claim, assesses whether a textual utterance affirms its truth ( $\oplus$ ), disputes it as false ( $\ominus$ ), or expresses a neutral or indeterminate position ( $\odot$ ).
- Truthfulness stance has the potential to be a useful tool in discerning how misinformation spreads and shapes decision-making in political discourse and health-related contexts.


**Claim: "California introduces new bill that would allow mothers to kill their babies up to 7 days after birth."**



Truthfulness Stance: **Positive**



Truthfulness Stance: **Neutral**



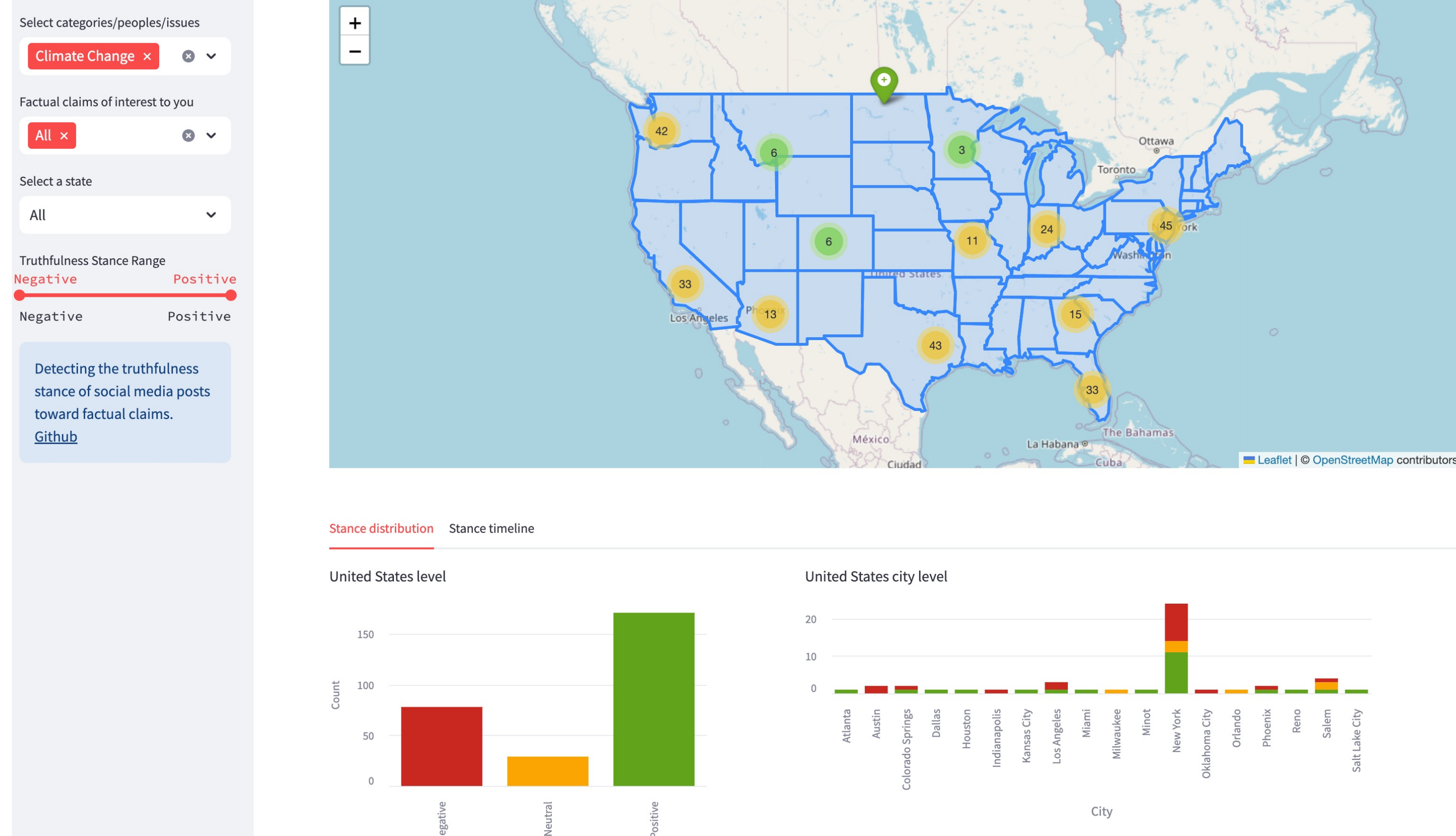
Truthfulness Stance: **Negative**



Truthfulness Stance: **No Stance**

## Real-World Applications

### Mapping Truthfulness Stance on Factual Claims for Geographical Analysis



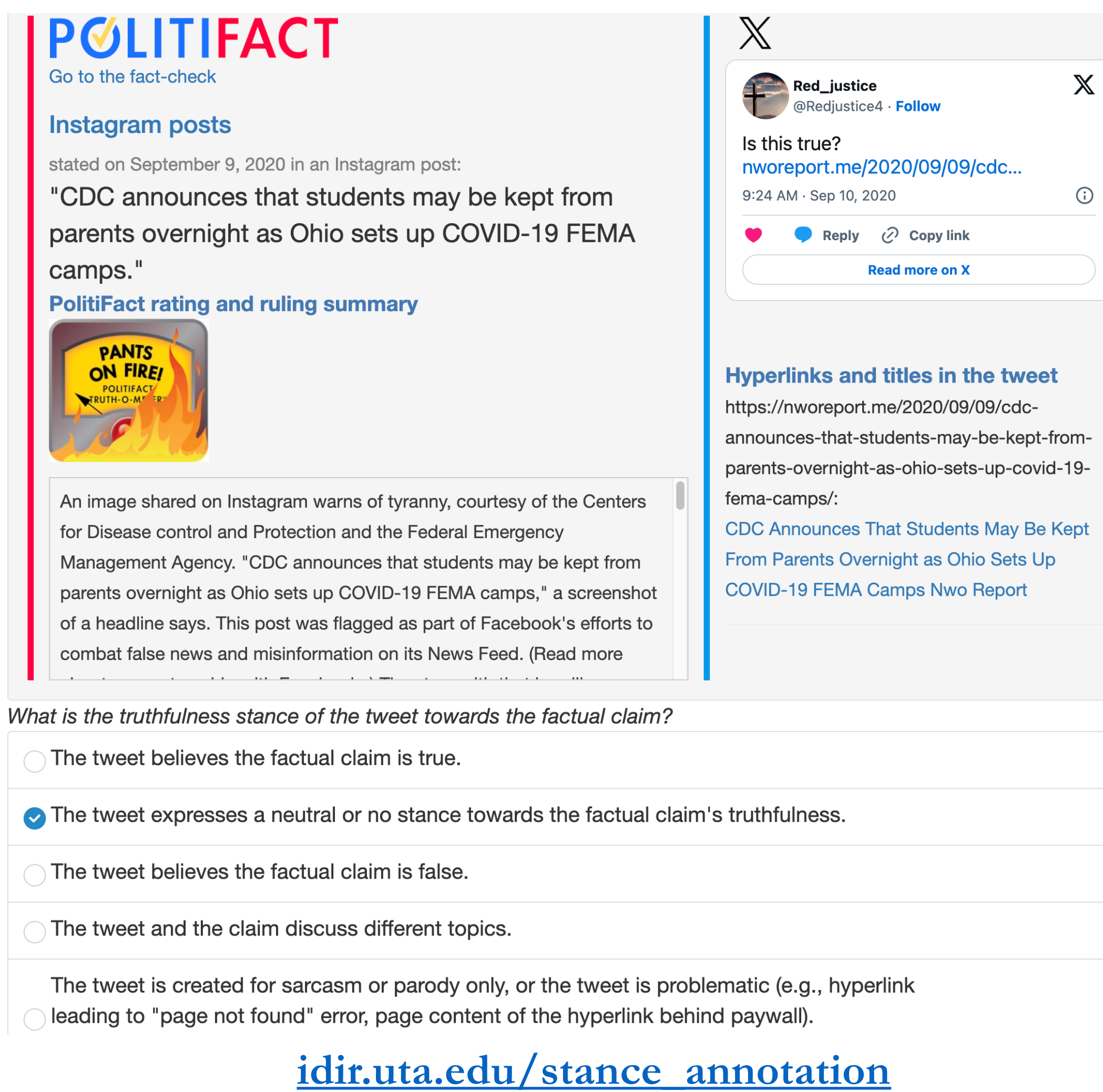
<https://trustmap.streamlit.app>

## TSD-CT Dataset

**Collecting Fact-checks:** Seven websites, 52,596 fact-checks (including associated factual claims) from 1995 to 2023.

**Collecting Claim-tweet Pairs:** Extracted keywords from factual claims and retrieved related tweets using Twitter API v2, resulting in 36,154 pairs. After sanitization, retained 5,793 pairs for human annotation.

**Data Annotation:** In-house annotation website with detailed instructions, a progress monitoring page, and a leaderboard. Collected 3,105 annotated pairs containing 1,520 unique claims.



What is the truthfulness stance of the tweet towards the factual claim?

☐ The tweet believes the factual claim is true.

☒ The tweet expresses a neutral or no stance towards the factual claim's truthfulness.

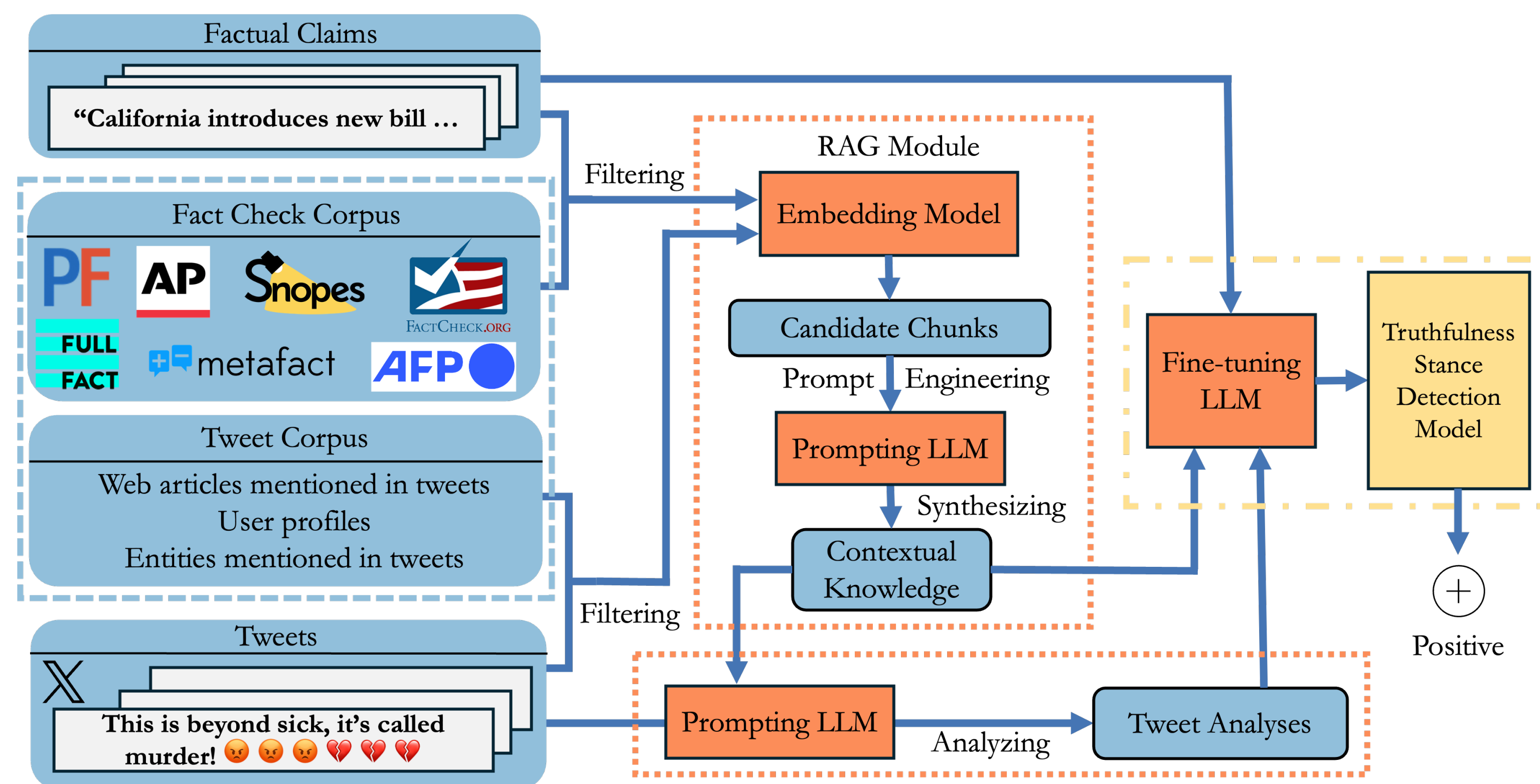
☐ The tweet believes the factual claim is false.

☐ The tweet and the claim discuss different topics.

☐ The tweet is created for sarcasm or parody only, or the tweet is problematic (e.g., hyperlink leading to "page not found" error, page content of the hyperlink behind paywall).

[idir.uta.edu/stance\\_annotation](http://idir.uta.edu/stance_annotation)

## The RATSD Framework



**Knowledge Corpora Construction:** ① Factual claim knowledge corpus encompasses 52,596 synthesized documents; ② Tweet knowledge corpus consists of 8,236 synthesized documents from 2010 to 2023.

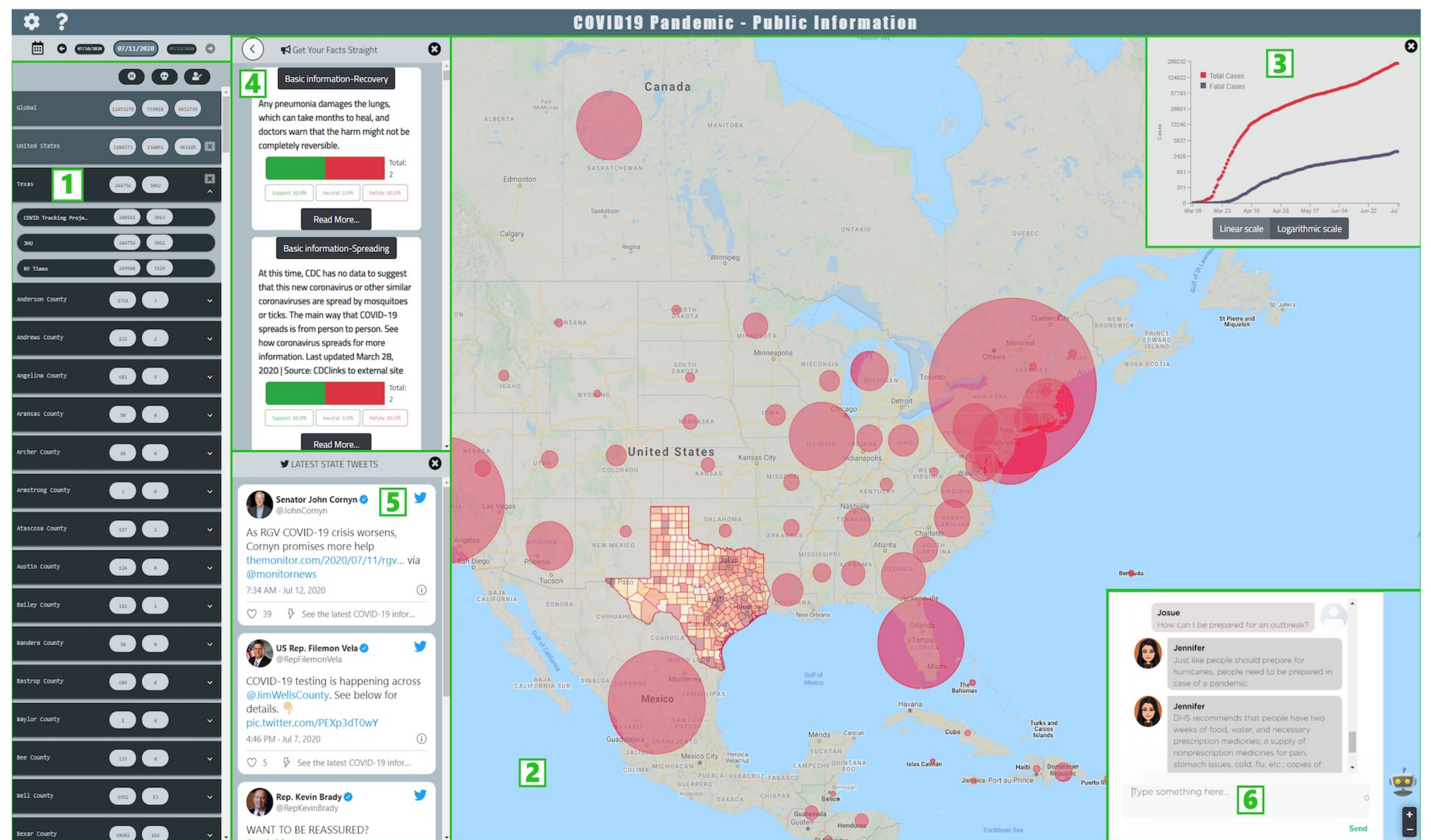
**Contextual Knowledge Generation:** (1) Document preprocessing (2) Relevant document selection (3) Relevant Chunk Retrieval (4) Prompting LLM

**Stance Analysis:** LLM is prompted using claim, tweet, contextual knowledge as the input to generate a narrative of tweet's truthfulness stance regarding claim.

**Classification Model:** LLM is fine-tuned using the claim-tweet pairs, stance analysis results, and contextual knowledge.

### A Dashboard for Mitigating the COVID-19 Misinfodemic

Components: 1) Geographic region selection panel 2) Interactive map 3) Timeline chart 4) Panel of facts 5) Government tweets 6) Chatbot



<https://idir.uta.edu/covid-19>

## Acknowledgements

This work is partially supported by the National Science Foundation award # 2346261. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that have contributed to the research results reported in this paper.



NAACL  
2025 SRW