

University of Thesis Proposal: Understanding Misinformation on Social Media

Through Truthfulness Stance



Zhengyuan Zhu

Dissertation Advisor: Chengkai Li

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.

TSD-CT Dataset

Collecting Fact-checks: Seven websites, 52,596 fact-checks (including associated factual

"CDC announces that students may be kept from

parents overnight as Ohio sets up COVID-19 FEMA

An image shared on Instagram warns of tyranny, courtesy of the Centers

for Disease control and Protection and the Federal Emergency

stated on September 9, 2020 in an Instagram post:

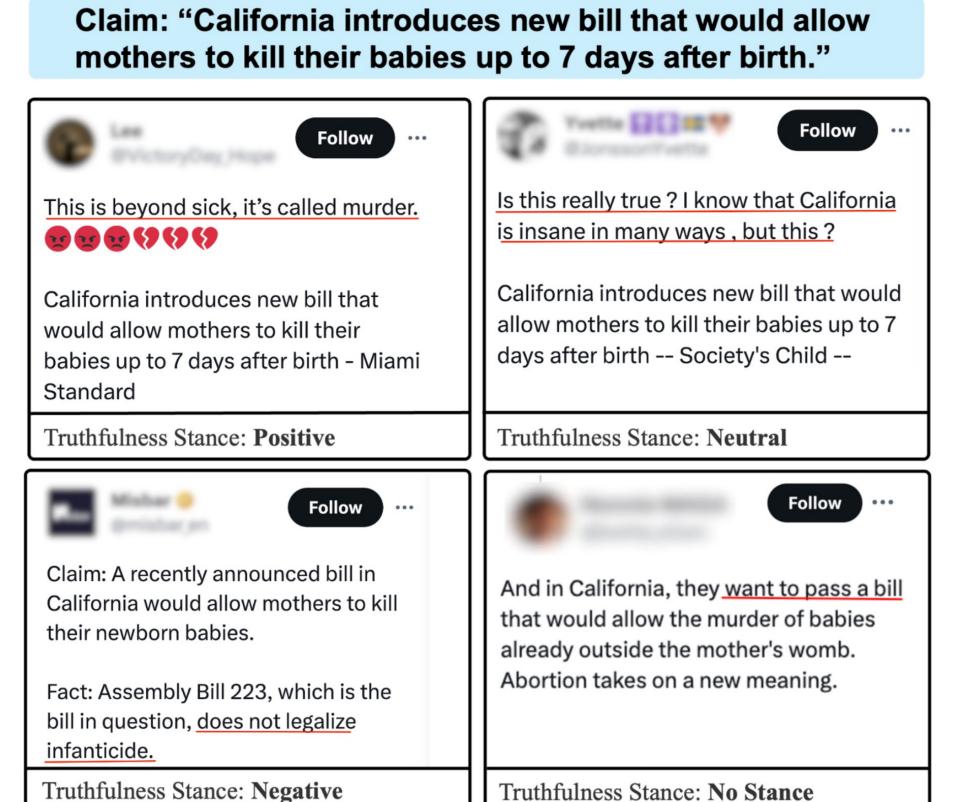
PolitiFact rating and ruling summary

PØLITIFACT

Instagram posts

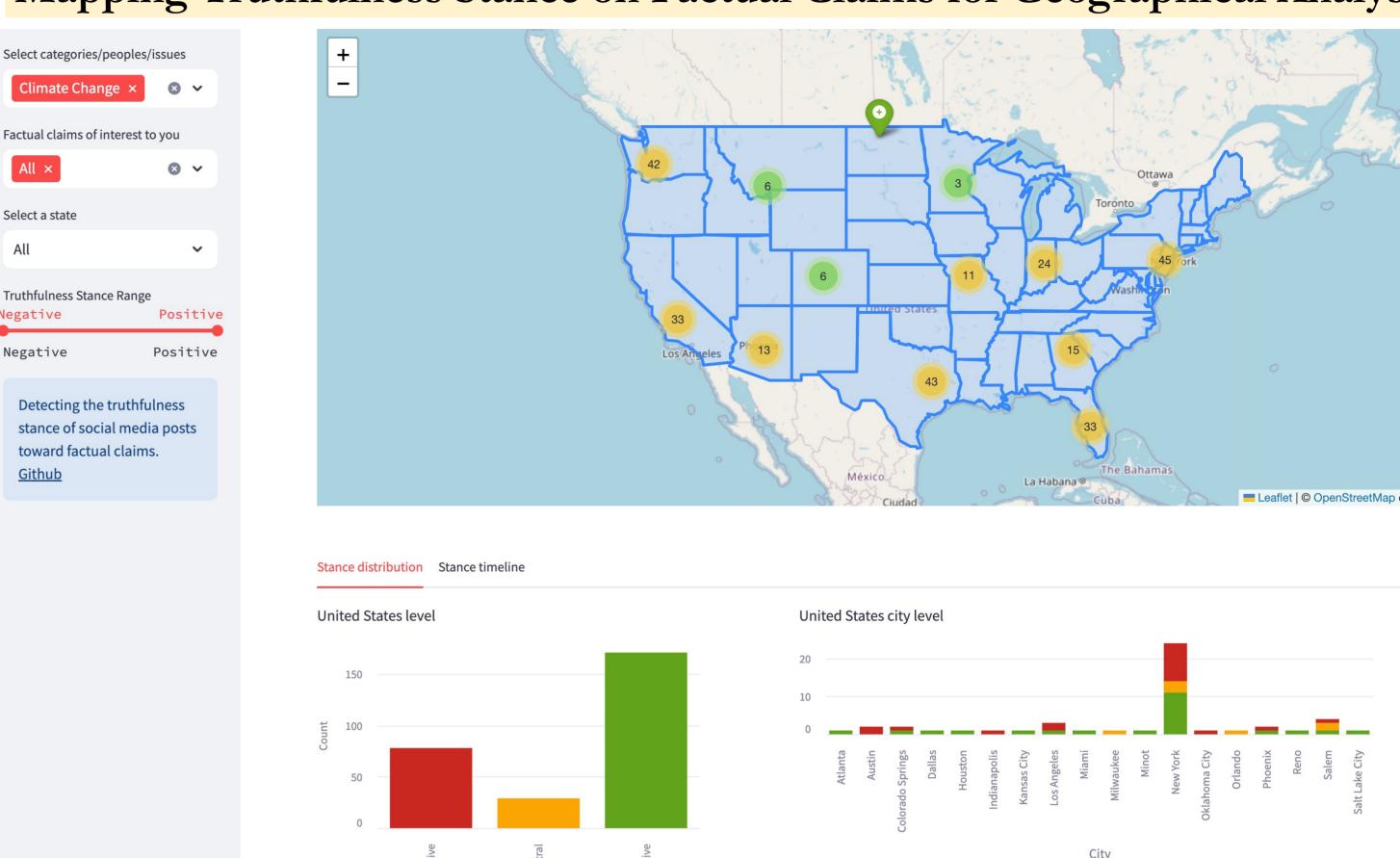
ON FIRE!

- Truthfulness stance: given a factual claim, assesses whether a textual utterance affirms its truth (\bigoplus) , disputes it as false (Θ) , 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 decisionmaking in political discourse and health-related contexts.





Mapping Truthfulness Stance on Factual Claims for Geographical Analysis



https://trustmap.streamlit.app

Topic	Truth-⊕	Truth-⊖	Misinfo-⊕	Misinfo-⊖	Accuracy	Macro F1
Public Health	77.2% (413)	22.8% (122)	53.8% (4,190)	46.2% (3,603)	61.7	39.3
Elections	40.0% (8)	60.8% (12)	58.9% (4,582)	41.1% (3,201)	40.6	29.3
Immigration	54.1% (53)	45.9% (45)	60.6% (3,023)	39.4% (1,969)	46.8	29.8
Economy	69.5% (534)	30.5% (234)	57.4% (1,174)	42.6% (872)	56.1	49.2
Abortion	79.9% (446)	20.1% (112)	44.3% (759)	55.7% (953)	67.8	59.6
Education	57.0% (231)	43.0% (174)	50.7% (849)	49.3% (827)	53.2	46.4
Crime	78.2% (223)	21.8% (62)	79.5% (1.030)	20.5% (265)	49.4	30.8
Environment	32.1% (131)	67.9% (277)	41.0% (661)	59.0% (953)	30.0	28.6
Political Leaning	Truth-⊕	Truth-⊖	Misinfo-⊕	Misinfo-⊖	Accuracy	Macro F1
Red States	67.9% (527)	32.1% (249)	63.7% (4,344)	36.3% (2,476)	52.1	35.3
Blue States	70.8% (659)	29.2% (272)	55.6% (4,259)	44.4% (3,402)	57.6	41.3
Swing States	73.4% (257)	26.6% (93)	62.5% (1,513)	37.5% (908)	55.5	38.7

Truthfulness stance distribution towards Truth and Misinformation across PolitiFact's topics. Truth-\(\oplus\) and Truth-\(\oplus\) denote positive and negative stances towards Truth, respectively. Misinfo- \bigoplus and Misinfo- \bigoplus denote positive and negative stances towards

announces-that-students-may-be-kept-fromparents-overnight-as-ohio-sets-up-covid-19-

fema-camps/

Red_justice
@Redjustice4 · Follow

Reply Copy link

Read more on X

Hyperlinks and titles in the tweet

CDC Announces That Students May Be Kept

From Parents Overnight as Ohio Sets Up

https://nworeport.me/2020/09/09/cdc-

9:24 AM · Sep 10, 2020

Management Agency. "CDC announces that students may be kept from COVID-19 FEMA Camps Nwo Report parents overnight as Ohio sets up COVID-19 FEMA camps," a screenshot of a headline says. This post was flagged as part of Facebook's efforts to Misinformation, respectively. What is the truthfulness stance of the tweet towards the factual claim's The tweet believes the factual claim is true.

Findings:

- The public struggles to distinguish between true and false claims. Public tends to believe claims are true regardless of their accuracy.
- People's judgments vary across topics. Higher belief in claims related to health. Low accuracy in topics such as election and environment.
- X users in blue states are better at distinguishing between true and false information compared to those in red states.

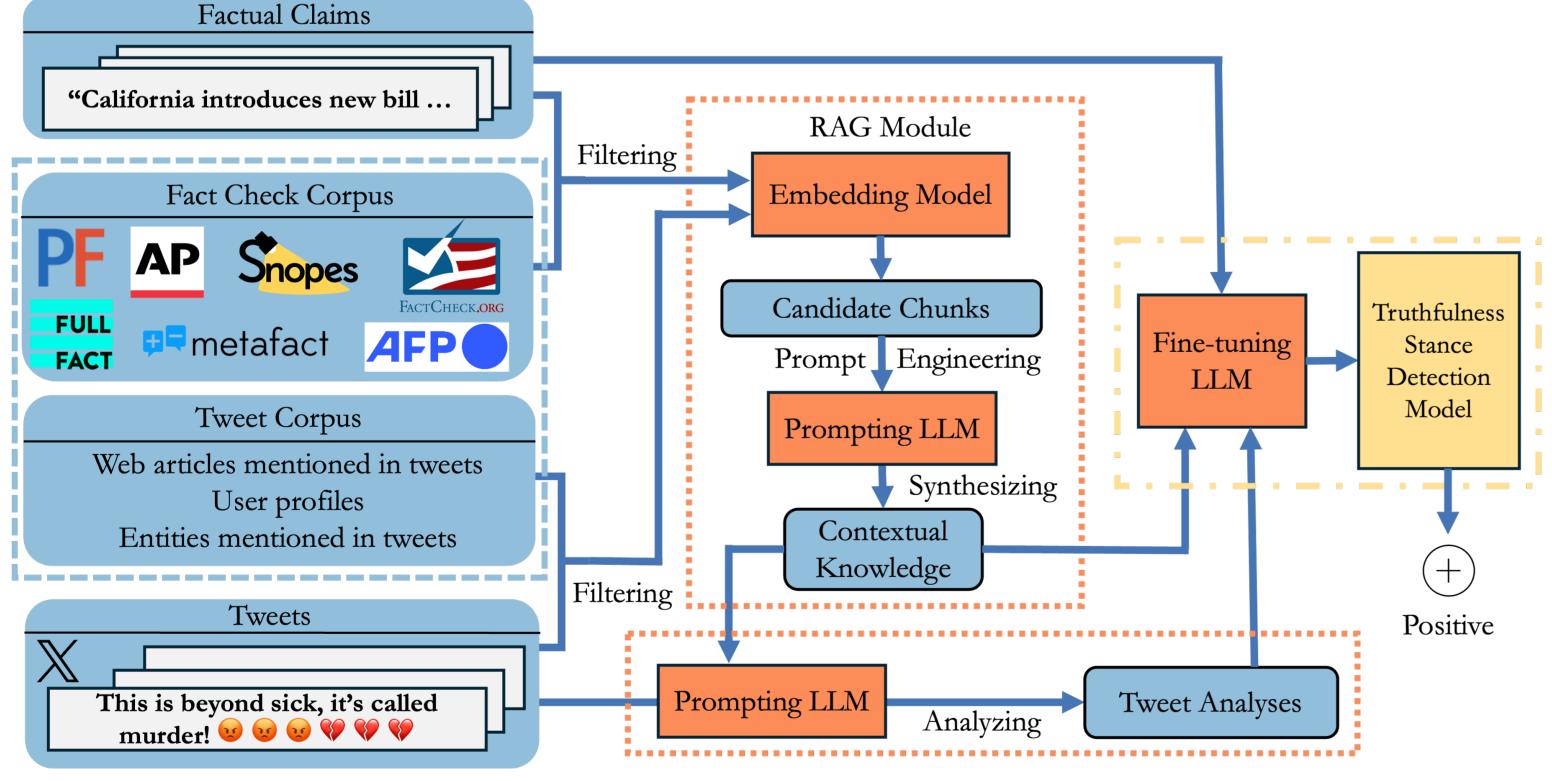
Collecting Claim-tweet Pairs: Extracted keywords from factual claims and retrieved related tweets using Twitter API v2, resulting in 36,154 pairs. After sanitization, retained 5,793 pairs for human annotation.

claims) from 1995 to 2023.

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

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

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.

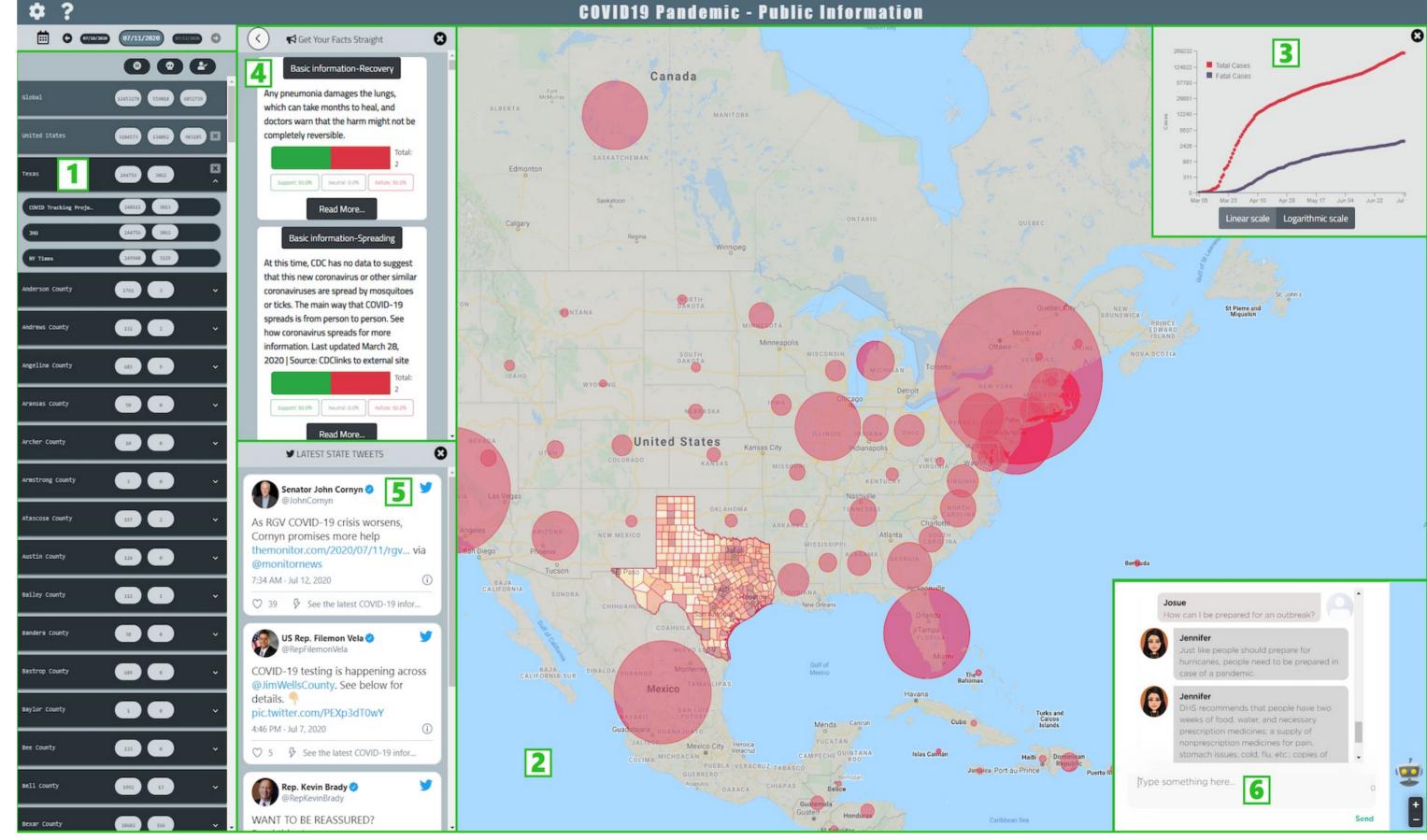
Contextual Knowledge Generation: (1) Document preprocessing (2) Relevant document selection (3) Relevant Chunk Retrieval (4) Prompting LLM

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.





