

On Mitigating Misinformation by Stance Detection

Annotation System: idir.uta.edu/wildfire_annotation/ Stance Detection API Interface: idir.uta.edu/stance_detection Covid-19 Dashboard: idir.uta.edu/covid-19

The Quest to Mitigate the Misinfodemic by Stance Detection

Social media posts make false claims

- “The COVID-19 vaccine is the **deadliest vaccine ever made**”
- “A viral photo shows the freedom convoy in Canberra, Australia, **with over 1.5 million**”
- “CNN confirms **Trump WILL be ARRESTED**”

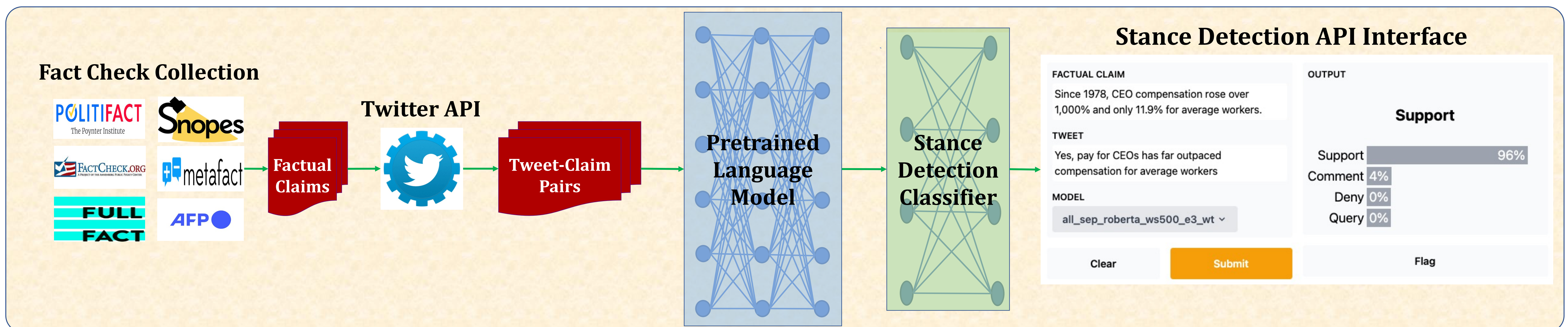
Misinformation exacerbates the vicious impact

- Drown out credible information
- Interfere with measures to contain the outbreak
- Deplete resources and overloads the health care system

How stance detection mitigates misinfodemic

- Stance is whether a tweet supports or refutes a factual claim
- Use stance detection for surveillance of misinfodemic and for discovering patterns of misinformation spreading

The Methodology of Stance Detection on Tweet-claim Pairs



Data Collection: Factual Claims, Tweets and Stance Dataset

Factual claims

- Built a daily fact check collection tool
- 52,805 factual claims from seven fact-checking websites

Tweets

- Used claim keywords and twitter academic research API to collect tweets.
- 65,011 distinct claim related tweets from 2007 to 2022

Stance detection annotation system

- Stance detection model trained using tweet-claim pairs annotated by human annotators

Performance of Stance Detection

Model	Dataset	Precision	Recall	MacroF1
BERT	Stance Annotation	0.48	0.50	0.47
Stance-BERT	Stance Annotation	0.80	0.78	0.78
Stance-BERT _{factcheck}	Stance Annotation	0.81	0.81	0.80
BERT	Politifact	0.32	0.35	0.31
Stance-BERT	Politifact	0.58	0.57	0.57
Stance-BERT _{factcheck}	Politifact	0.57	0.57	0.56
SiamNet [Conforti et al. 2020]	WTWT2020	-	-	0.622
BERT	WTWT2020	0.84	0.85	0.84
Stance-BERT	WTWT2020	0.85	0.85	0.85
Stance-BERT _{factcheck}	WTWT2020	0.86	0.86	0.86

Case Study: COVID-19 Dashboard

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

Factual claim: America is one of the highest-taxed nations in the world

Tweet: Trump says America is one of the highest-taxed nations in the world. That's bullshit. <http://53eig.ht/2a4BsT8>

More Context

What is the truthfulness stance of tweet towards factual claim?

☐ Tweet author believes the veracity of the factual claim

☐ Tweet author cannot confirm the veracity of the factual claim

☐ Tweet author doesn't believe the veracity of the factual claim

☐ Tweet discusses a topic that is unrelated to the factual claim

☐ Tweet-claim pair is problematic: Tweet is non-English/incomprehensive or claim is a noun phrase/question.

Submit Skip this sentence Modify My Previous Responses

