Structured Understanding of Social Media Discourse: An Automated Framework for Taxonomy Construction

THE UNIVERSITY OF TEXAS AT ARLINGTON

NAACL Student Research Workshop (SRW) 2025

Haiqi Zhang

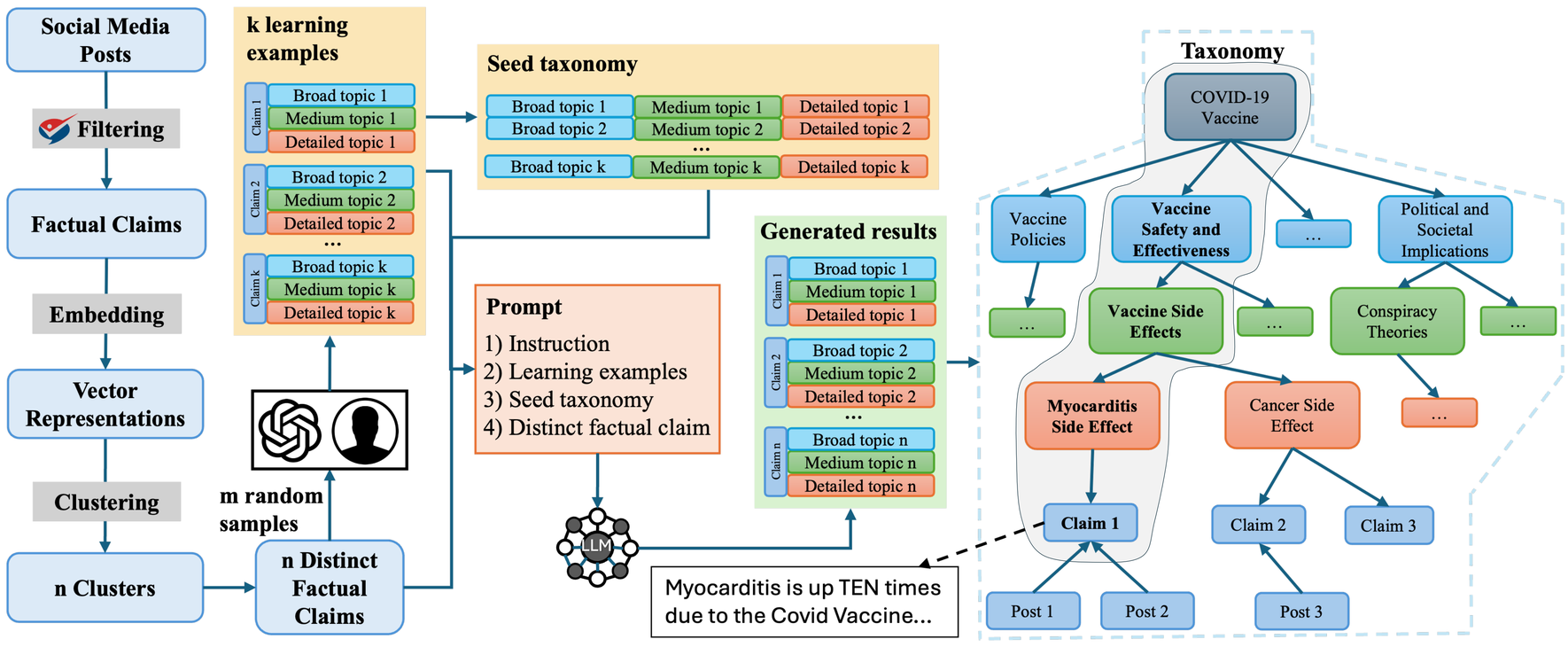
Dissertation Supervisor: Chengkai Li

Innovative Data Intelligence Research Laboratory

Introduction

- Understanding online narratives and public perception, especially regarding misinformation, is challenging.
- Need scalable, automated tools to organize and analyze the large volume and variety of social media posts.
- Introduce a novel LLM-based framework, LLMTaxo, for constructing multi-level taxonomy of factual claims.
- First taxonomy categorizing factual claims in a specific topic domain at broad, medium, and detailed levels.
- New evaluation metrics to assess taxonomy quality and claim-topic alignment. Validated across diverse datasets.

LLMTaxo Framework

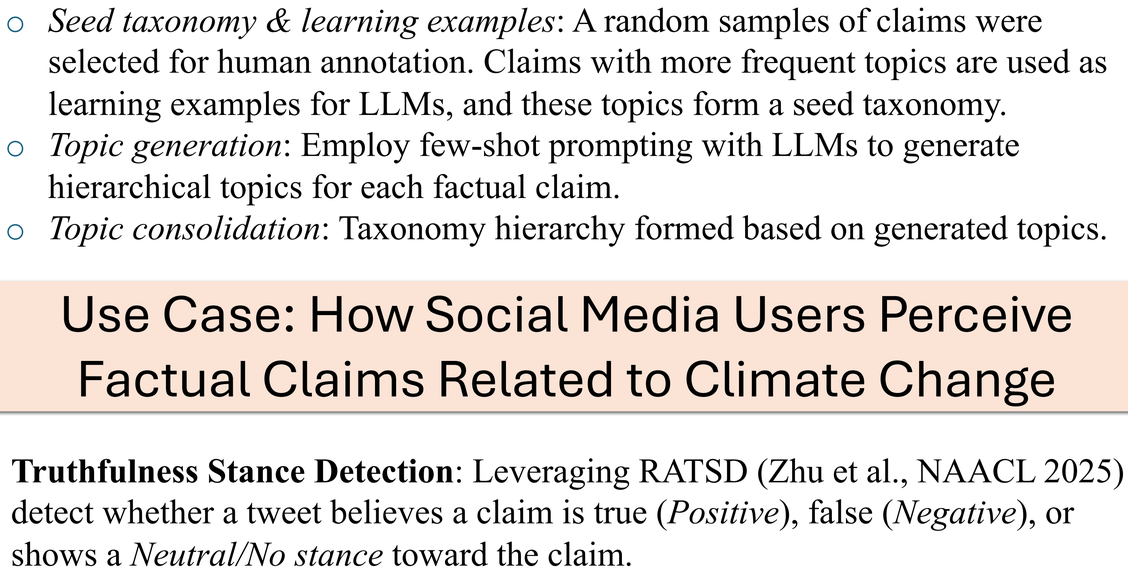


Methodology

- **Datasets**:
  - *COVID-19 Vaccine* (CV): 384,676 tweets, 01/23-04/23.
  - *Climate Change* (CC): 229,913 Facebook posts, 01/24-05/24.
  - *Cybersecurity* (CS): 107,905 Facebook posts, 01/24-05/24.
- **Claim Detection**: Apply ClaimBuster (Hassan et al., 2017) to identify check-worthy factual claims and retain claims with a score above 0.5.
- **Distinct Claim Identification**: Employ HDBSCAN (Campello et al., 2013) clustering and Sentence-BERT (Reimers et al., 2019) to group semantically similar posts and reduce redundancy.
- **Taxonomy Construction**:
  - *Seed taxonomy & learning examples*: A random samples of claims were selected for human annotation. Claims with more frequent topics are used as learning examples for LLMs, and these topics form a seed taxonomy.
  - *Topic generation*: Employ few-shot prompting with LLMs to generate hierarchical topics for each factual claim.
  - *Topic consolidation*: Taxonomy hierarchy formed based on generated topics.

Use Case: How Social Media Users Perceive Factual Claims Related to Climate Change

- **Truthfulness Stance Detection**: Leveraging RATSD (Zhu et al., NAACL 2025) to detect whether a tweet believes a claim is true (*Positive*), false (*Negative*), or shows a *Neutral/No stance* toward the claim.
- **Claim-tweet Pairs**:
  - *Factual claims (729)*: Collected from five fact-checking websites using climate change-related keywords, such as "global warming" and "greenhouse gas."
  - *Tweets (13,050)*: Using tokens extracted from factual claims (nouns, verbs, adjectives) to collect tweets from ***X***.



Stance Distribution Towards Truth and Misinformation Across Broad Topics

Broad Topic	Truth-⊕	Truth-⊖	Misi-⊕	Misi-⊖	Accuracy	Macro F1
Climate Science	81.7% (524)	18.3% (117)	72.5% (377)	27.5% (143)	0.575	0.524
Economy	70.5% (146)	29.5% (61)	72.5% (351)	27.5% (133)	0.404	0.404
Energy	82.2% (264)	17.8% (57)	74.7% (124)	25.3% (42)	**0.628**	**0.530**
Environment	77.5% (533)	22.5% (155)	74.4% (1040)	25.6% (357)	0.427	0.423
Government Policies	83.2% (183)	16.8% (37)	**69.5%** (205)	**30.5%** (90)	0.530	0.514
Health	**88.7%** (180)	**11.3%** (23)	**77.9%** (169)	**22.1%** (48)	0.543	0.493
Politics	**69%** (363)	**31%** (163)	75.7% (1635)	24.3% (525)	**0.331**	**0.329**
Technology	74.8% (86)	25.2% (29)	69.8% (120)	30.2% (52)	0.481	0.473

Quality Evaluation

- **Evaluators**: Three humans and GPT-4.
- **Evaluation Metrics**:
  - Taxonomy quality: clarity (M_1), hierarchical coherence (M_2), orthogonality (M_3), completeness (M_4).
  - Claim-detailed topic alignment: accuracy (Ac), granularity (Gr).
- The results affirm the overall efficacy of our framework, with taxonomies generally receiving high ratings.
- GPT-4o mini consistently outperformed Zephyr in taxonomy evaluation (M_1, M_2, M_3, M_4), although Zephyr has better claim-topic alignment on CV.

Dataset	Evaluator	Model	M_1	M_2	M_3	M_4	Ac	Gr
CV	*Human*	Zephyr	4.2	4.0	3.4	4.3	4.3	4.1
		GPT-4o mini	4.3	4.2	3.8	4.7	3.8	3.8
	GPT-4	Zephyr	3.5	3.7	3.0	3.7	3.9	4.6
		GPT-4o mini	4.3	4.3	4.0	4.7	4.0	4.4
CC	*Human*	Zephyr	4.6	4.2	3.9	4.6	3.4	4.6
		GPT-4o mini	4.7	4.7	4.2	4.8	4.0	4.0
	GPT-4	Zephyr	3.5	3.3	4.0	4.0	3.3	3.4
		GPT-4o mini	4.3	4.0	4.5	4.3	3.7	4.1
CS	*Human*	Zephyr	4.3	3.9	3.7	4.4	3.9	4.0
		GPT-4o mini	4.5	4.0	4.0	4.6	4.4	4.2
	GPT-4	Zephyr	3.3	3.0	3.0	3.3	3.8	4.3
		GPT-4o mini	4.3	4.7	3.5	4.7	4.1	4.9

Use Case Sample Results

Claim	Tweet	Truthfulness Stance	Broad Topic	Medium Topic	Detailed Topic
Air pollution linked to greater risk of dementia.	People over 50 in areas with the highest levels of nitrogen oxide in the air showed a 40% greater risk of developing dementia than those with the least NOx #airpollution.	⊕	Health	Air Pollution	Impacts on Brain Health
Sen. Lindsey Graham supports the Green New Deal.	Facebook removed an ad by Adriel Hampton showing Sen. Lindsey Graham backing the Green New Deal.	⊙	Politics	Climate Change Advocacy	Politicians ' Stance
The Earth is warming because of the sun's changing distance from the Earth, not because of carbon emissions.	Enough with your pseudo-scientific. Actual science has proven the relationship to human carbon emissions and not cycles of sun /earth distance.	⊖	Climate Science	Climate Feedback Mechanisms	Misconceptions

Conclusion

- The proposed framework can effectively automate taxonomy construction.
- Evaluation across topic domains highlights its potential for broader applications.
- The use case shows one of the taxonomy applications. People's judgments vary across topics. Higher belief in claims related to health. Lower accuracy on topics like politics, economy, and environment.

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