

Telling a Lie: Analyzing the Language of Information and Misinformation during **Global Health Events**



Ankit Aich and Natalie Parde

Department of Computer Science

University of Illinois at Chicago

{aaich2, parde}@uic.edu

ankitaich.com, natalieparde.com

Case Studies These are the things and events that happen when a sacred landslide election victory is so unceremoniously & viciously stripped away from great patriots who have been badly & unfairly treated for so long. Go home with love & in peace. Remember this day forever 11:01 PM · Jan 6, 2021 NDIA-Science behind 5th April 2020 lamp ritual Scientific reason behind 5th April 2020 to light lamps , diyas candles at 9 pm for 9 minutes . Number 9 represents mars , number 5 mercury and 2 represents moor Jumber 9 represents planet Mars . At present mars is in makar rashi with planet Jupiter and Saturn. Moon is exactly situated at 180 egrees from these 3 planets he conjunctions of moon and mars will create an atmosphe park at 9.00 pm on 5th April 2020 that will increase the magnet eld of planets earth. Plus with lamps lightened and electric light closed. It will increase the power of magnetic field plus moon wil be at the shortest distance from earth. Moon is considered to control mind and water in human body It will increase the human immunity in the Indian subcontinen it will produce oxygen from lamps though in small quantities from each lamp but if we consider 50 percent of Indian population lightin lamps imagine the oxygen level in environment plus the magneti field will use this oxygen reduce the impact of corona virus from atmosphere and help in respiratory immunity of human body and This combination of moon and mars help nature's fight against pandemic covid 19

Motivations

- Widespread misinformation during the COVID-19 pandemic
- Minimal data from print or social media regarding previous health events





Observed differences between information and misinformation

What is misinformation?



be an innocen

Misinformation: Might be a lack of intent (e.g., if spread due to genuine concern), but still harmful and can cause widespread effects.

Disinformation: Requires intent (e.g., is caused as a result of someone's agenda), and almost always has harmful side effects.



Data Sources









- Sources were a mix of print media, social media, online media and communications from health institutions
- Articles ranged in publication date from 1900 to 2021
- Articles included an average of 330 words and 6800 characters

Data Annotation

Ë

111

SOCIAL MEDIA ANNOTATION

NEWS ANNOTATION



Readability Index	$RI = 4.71 \frac{c}{w} + 0.5 \frac{w}{s} - 21.43$	
Ease of Reading	$ER = 206.835 - 1.015 \frac{w}{s} - 84.6 \frac{l}{w}$	
Difficult Words	N(w) ∀ w ∈ <i>TextStat Easy words</i> or w has ≥ 2 syllables	
+NRC Scores	Number of Positive Words in NRC Word Emolex	
-NRCScores	Number of Negative Words in NRC Emolex	

C = string length, W = # words, S = # sentences in corpus, I = # syllables





- First misinformation dataset spanning all major health events • In depth analysis shows recurring themes and features across misinformation and factual information
 - Personal pronouns are more prevalent in misinformation
 - Analytical complexity is lower in misinformation
 - Lexicon-based trust features correlate strongly with misinformation class
 - Politically polarized n-grams occur more frequently with misinformation
- Proof-of-concept modeling experiment establishes strong performance at distinguishing between factual articles and misinformation in this dataset • $F_1 = 0.7$
 - Accuracy = 75%

• Can be used to train systems for future fine-grained misinformation detection

Check out our paper for more details!

