# Vaccine sentiments and hesitancy on social media: A machine learning-powered, real-time monitoring system

Jingcheng Du[1]; Amanda Eiden[2]; Liang-Chin Huang[1]; Long He[1]; Siwei Wang[1]; Jingqi Wang[1]; Frank J. Manion[1]; Xiaoyan Wang[1]; Lixia Yao[2]

[1]Melax Tech, Houston, TX, USA;
[2]Merck & Co., Inc., Rahway, NJ, USA

## Social media listening and vaccine hesitancy

- Rapid social media growth has facilitated the spread of mis- and disinformation on vaccines and contributed to negative public sentiments towards vaccination
- Understanding vaccine hesitancy is crucial for developing effective interventions, public health education, and vaccination promotion strategies
- Objective: To leverage natural language processing (NLP) algorithms to monitor and analyze vaccine sentiment and hesitancy across three social media platforms

### Survey disadvantages

- Expensive, time-consuming, and labor-intensive
- Challenging to recruit a broad range of participants
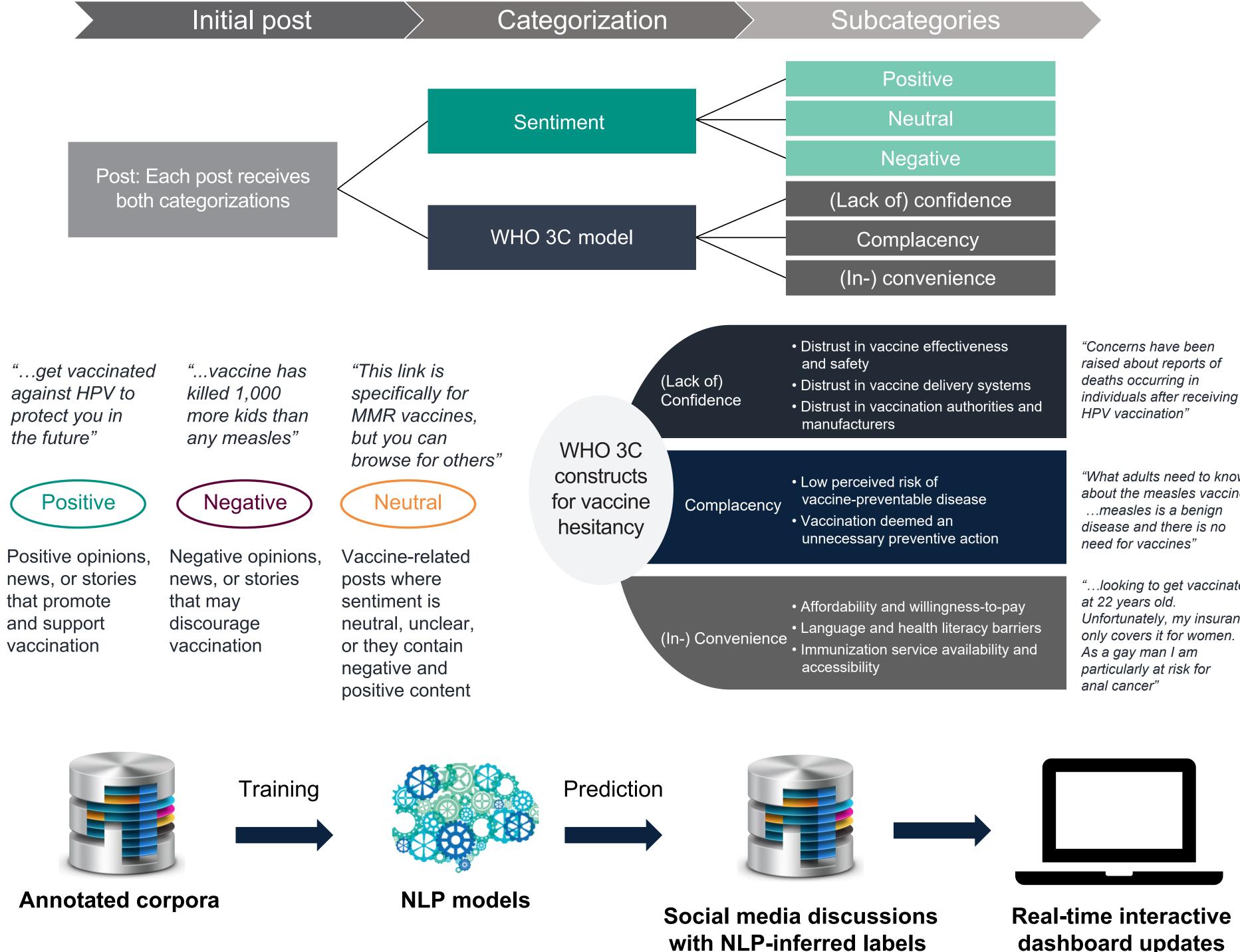- Difficult to track long-term changes

### Social media + NLP advantages

- Large volume of information from broad population
- Voluntarily contributed by users in the current moment or in need
- Easy to stay up-to-date

## Data collection and NLP for data labeling

- We collected and analyzed social media discussions from 2011-01-01 to 2021-10-31 related to human papillomavirus (HPV) vaccines; measles, mumps, and rubella (MMR) vaccines; and general, unspecified vaccines, from Twitter, Reddit, and YouTube
- NLP algorithms classify vaccine sentiment as positive, neutral, or negative and further predict vaccine hesitancy aligning with the World Health Organization's 3C vaccine hesitancy framework (Complacency, Confidence, and Convenience)
- We manually curated a benchmark data set that annotated 30,000 Twitter tweets, 15,000 Reddit posts, and 15,000 YouTube comments

### Figure 1. Vaccine sentiment and hesitancy information model



## Real-time interactive dashboard

- User-friendly, web-based visualization dashboard that displays trends in vaccine sentiment and hesitancy over time and across geographies
- Allows for comparisons of sentiment and hesitancy across social media platforms and vaccine targets

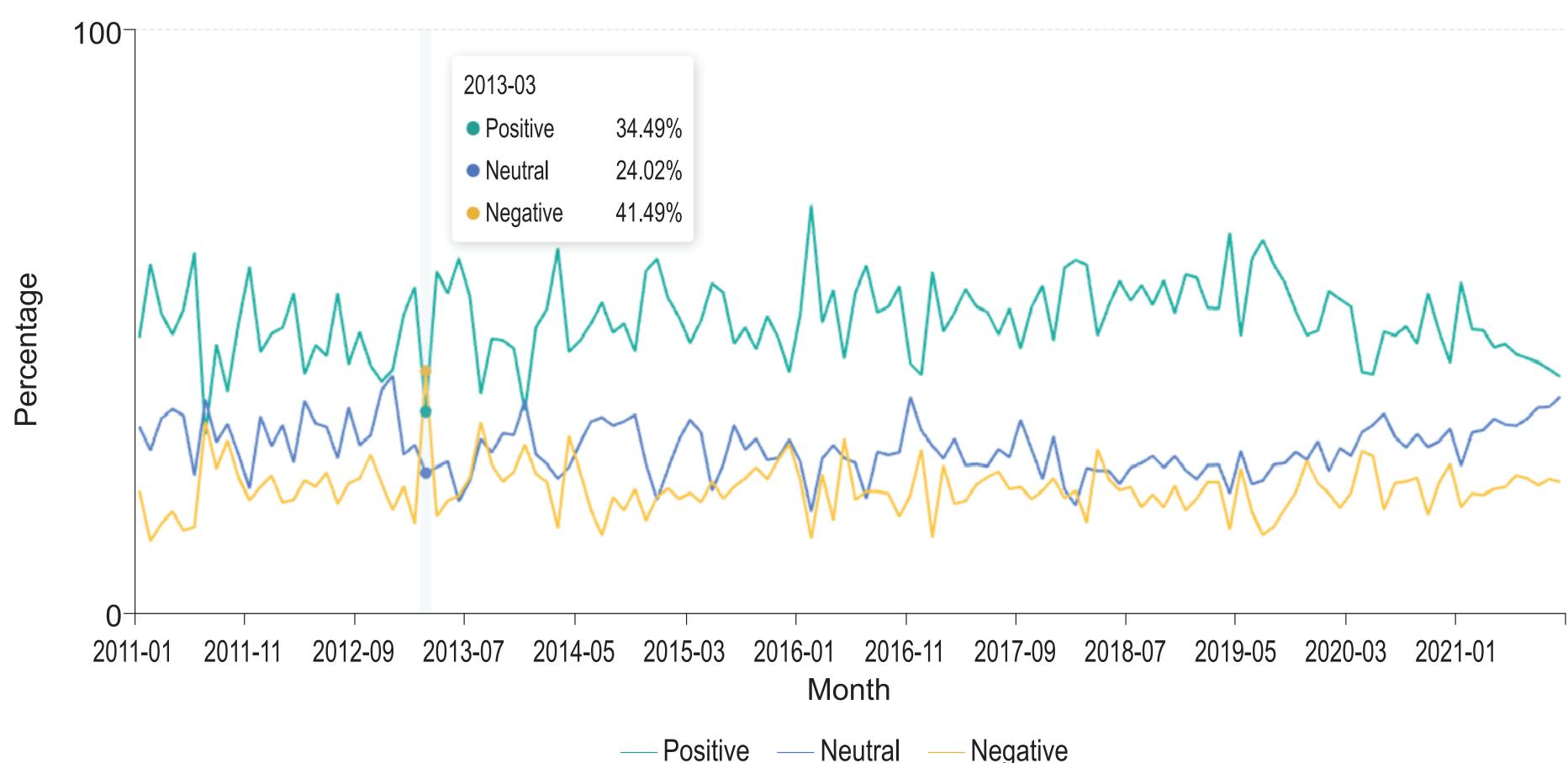### Figure 2. Selected dashboard views



(**A**) Overview of social media vaccine-related discussions – descriptive statistics; (**B**) Temporal trends to demonstrate variations across time – users can add data filters for specific vaccine and sentiment topics; (**C**) Geo-clustering of social media discussions – both US state-level and world view are provided; (**D**) comparison view that enables users to select vaccine types, social media data sources, sentiment and hesitancy topics, and geo-locations.

## NLP results and selected case studies

- >90 million social media discussions were collected. Machine learning-based NLP algorithms achieved prediction accuracy scores ranging from 0.51 to 0.78 on vaccine sentiment and 0.69 to 0.91 on vaccine hesitancy
- Temporal trends revealed variations over time in social media activity across vaccine categories regarding vaccine sentiments and hesitancy; for example, sentiments on Twitter for HPV vaccine trended generally more positively than neutrally and negatively over time

### Figure 3. Aggregation of three social media platform data sources to evaluate vaccine sentiment for HPV vaccine-related posts



- March 2013, negative sentiment posts had a higher ratio (41.49%) than positive (34.49%) or neutral (24.02%) posts on all three platforms
- Two articles from March 2013, titled "Worried Parents Balk at HPV Vaccine for Daughters" by National Public Radio and "Side Effect Fears Stop Parents from Getting HPV Vaccine for Daughters" by CBS News, may be associated with the negative sentiment expressed on social media

### Contact:
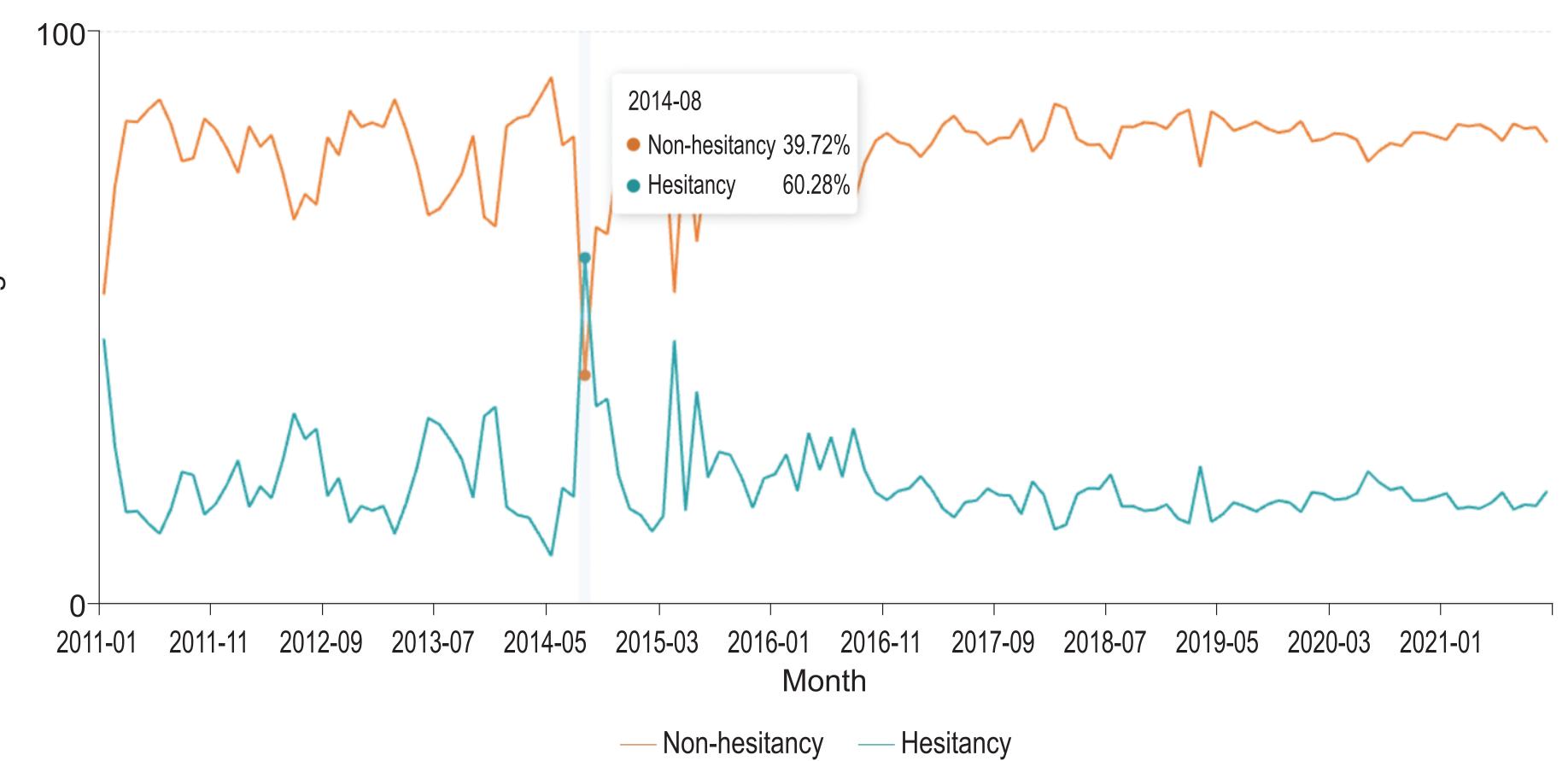
Jingcheng.Du@melaxtech.com and Amanda.Eiden@merck.com

### Figure 4. Comparison of vaccine hesitancy for MMR vaccine-related posts



- In August 2014, higher vaccine hesitancy for MMR was observed across three social media platforms
- Two news articles were posted that month on the link between MMR vaccines and autism, which may have been associated with this peak in hesitancy posts
- The articles were titled "Journal Questions Validity of Autism and Vaccine Study" by CNN and "Whistleblower Claims CDC Covered Up Data Showing Vaccine-Autism Link" by TIME magazine

### Limitation

- Generalizability may be limited due to the population bias for different social media platforms and the language restriction to English only
- The misclassification rates of NLP models may have influenced the results

### Conclusions

- Our study demonstrates that real-time temporal and geographical analyses that generate actionable insights to improve vaccine uptake is possible
- When an increase in vaccine hesitancy posts is detected, such a tool could be used to guide public health professionals to increase communications and ensure the public receives accurate information