The Role of the Crowd in Countering Misinformation: A Case Study of the COVID-19 Infodemic

Nicholas Micallef* [1], Bing He* [2], Srijan Kumar [2], Mustaque Ahamad [2], Nasir Memon [3]
[1] New York University Abu Dhabi, [2] Georgia Institute of Technology, [3] New York University
[1]nicholas.micallef@nyu.edu, [2]{bhe46,srijan,mustaq}@gatech.edu, [3] memon@nyu.edu

2020 IEEE International Conference on Big Data (Big Data) | 978-1-7281-6251-5/20/$31.00 ©2020 IEEE | DOI: 10.1109/BigData50022.2020.9377956

***Abstract*—Fact checking by professionals is viewed as a vital defense in the fight against misinformation. While fact checking is important and its impact has been significant, fact checks could have limited visibility and may not reach the intended audience, such as those deeply embedded in polarized communities. Concerned citizens (i.e., the crowd), who are users of the platforms where misinformation appears, can play a crucial role in disseminating fact-checking information and in countering the spread of misinformation. To explore if this is the case, we conduct a data-driven study of misinformation on the Twitter platform, focusing on tweets related to the COVID-19 pandemic, analyzing the spread of misinformation, professional fact checks, and the crowds response to popular misleading claims about COVID-19.**

In this work, we curate a dataset of false claims and statements that seek to challenge or refute them. We train a classifier to create a novel dataset of 155,468 COVID-19-related tweets, containing 33,237 false claims and 33,413 refuting arguments. Our findings show that professional fact-checking tweets have limited volume and reach. In contrast, we observe that the surge in misinformation tweets results in a quick response and a corresponding increase in tweets that refute such misinformation. More importantly, we find contrasting differences in the way the crowd refutes tweets, some tweets appear to be opinions, while others contain concrete evidence, such as a link to a reputed source. Our work provides insights into how misinformation is organically countered in social platforms by some of their users and the role they play in amplifying professional fact checks. These insights could lead to development of tools and mechanisms that can empower concerned citizens in combating misinformation. The code and data can be found in this link.[1]

***Index Terms*—Misinformation, Counter-misinformation, Social Media, Dataset**

I. Introduction

The world-wide spread of COVID-19 has led to considerable amount of related misinformation on the web and the social media ecosystem. WHO has termed the situation as a global infodemic [1]. As social media platforms become a primary means to acquire and exchange news and information during crisis times such as COVID-19 [2]–[8], the lack of clear distinction between true and false information can be dangerous. Some of the false claims related to COVID-19 have already had severe harmful consequences, including violence [9] and over 800 deaths [10]. Thus, combating the spread of false information is of critical importance.

Professional fact checkers can play an important role in controlling the spread of misinformation on online platforms [11]. During the COVID-19 infodemic, the International Fact Checking Network (IFCN) verified over 6,800 false claims related to the pandemic until May 20, 2020. Social media platforms use these fact checks to flag and sometimes remove misinformation content. However, false information still prevails on social platforms because the ability of fact checking organizations to use social media to disseminate their work can be limited [12]. For example, on Facebook, content from the top 10 websites spreading health misinformation had almost four times as many estimated views as equivalent content from reputable organizations (e.g., CDC, WHO).[2]

In addition to professional fact checkers, ordinary citizens, who are concerned about misinformation, can play a crucial role in organically curbing its spread and impact. Compared to professional fact checkers, concerned citizens, who are users of the platform where misinformation appears, have the ability to directly engage with people who propagate false claims either because of ignorance or for a malicious purpose. They can back up their arguments using professional fact checks and trusted sources, whenever available. The cohort of ordinary citizens is also commonly referred to as *crowd*. Thus, the role of crowd or citizens who are concerned about misinformation can be critically important. The goal of this work is to study the nature and extent of the role that concerned citizens play in responding to misinformation.

We use a broad definition of *misinformation* which includes falsehoods, inaccuracies, rumors, decontextualized truths, or misleading leaps of logic, all regardless of the intention of the spreader [13], [14]. In this work, we focus on COVID-19 related misinformation on Twitter and utilize a data-driven approach to investigate how fact checks and other organic user responses attempt to refute and counter it. We explore two popular misinformation topics: *fake cures* and *5G conspiracy*

* co-first authors. The first two authors contributed equally to this work.
[1]http://claws.cc.gatech.edu/covid_counter_misinformation.html

[2]https://secure.avaaz.org/campaign/en/facebook_threat_health/

Characterizing and Predicting Social Correction on Twitter

Yingchen Ma
Georgia Institute of Technology
Atlanta, Georgia, USA
yma473@gatech.edu

Bing He
Georgia Institute of Technology
Atlanta, Georgia, USA
bhe46@gatech.edu

Nathan Subrahmanian
Brandeis University
Waltham, Massachusetts, USA
nsubrahmanian@brandeis.edu

Srijan Kumar
Georgia Institute of Technology
Atlanta, Georgia, USA
srijan@gatech.edu

ABSTRACT

Online misinformation has been a serious threat to public health and society. Social media users are known to reply to misinformation posts with counter-misinformation messages, which have been shown to be effective in curbing the spread of misinformation. This is called social correction. However, the characteristics of tweets that attract social correction versus those that do not remain unknown. To close the gap, we focus on answering the following two research questions: (1) "Given a tweet, will it be countered by other users?", and (2) "If yes, what will be the magnitude of countering it?". This exploration will help develop mechanisms to guide users' misinformation correction efforts and to measure disparity across users who get corrected. In this work, we first create a novel dataset with 690,047 pairs of misinformation tweets and counter-misinformation replies. Then, stratified analysis of tweet linguistic and engagement features as well as tweet posters' user attributes are conducted to illustrate the factors that are significant in determining whether a tweet will get countered. Finally, predictive classifiers are created to predict the likelihood of a misinformation tweet to get countered and the degree to which that tweet will be countered. The code and data is accessible on https://github.com/claws-lab/social-correction-twitter.

CCS CONCEPTS

• **Information systems** → Social networks.

KEYWORDS

Misinformation, Counter-misinformation, Social Correction, Twitter, COVID-19 vaccines

ACM Reference Format:
Yingchen Ma, Bing He, Nathan Subrahmanian, and Srijan Kumar. 2023. Characterizing and Predicting Social Correction on Twitter. In *15th ACM Web Science Conference 2023 (WebSci '23), April 30–May 01, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3578503.3583610

WebSci '23, April 30–May 01, 2023, Austin, TX, USA

ACM ISBN 979-8-4007-0089-7/23/04.
https://doi.org/10.1145/3578503.3583610

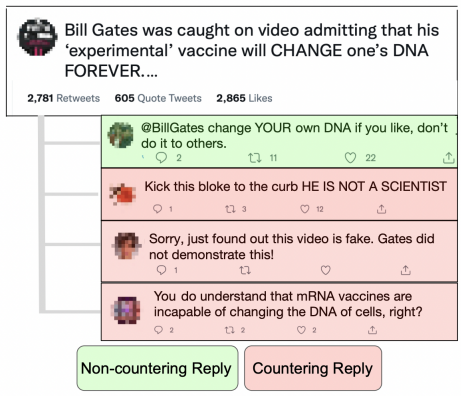

Figure 1: Examples of misinformation tweets and counter-misinformation replies.

1 INTRODUCTION

Online misinformation leads to societal harm including diminishing trust in vaccines and health policies [6, 49], damaging the well-being of users consuming misinformation [35, 63], encouraging violence and harassment [5, 60], and posing a danger to democratic processes and elections [57–59]. The problem has been exacerbated during the COVID-19 pandemic [40, 56]; particularly, COVID-19 vaccine misinformation including false claims that the vaccine causes infertility, contains microchips, and even changes one's DNA and genes has fueled vaccine hesitancy and reduced vaccine uptake [56]. Therefore, it is crucial to restrain the spread of online misinformation [36, 40]. In this work, we use a broad definition of misinformation which contains rumors, falsehoods, inaccuracies, decontextualized truths, or misleading leaps of logic [35, 68].

To combat misinformation, various countermeasures have been developed [40, 42, 65]. Recent work has shown that ordinary users of online platforms play a crucial role in countering misinformation. According to the research study by Micallef et al. [40], the vast majority (96%) of online counter-misinformation responses are made by ordinary users, with the remainder being made by professionals such as fact-checkers and journalists. While fact-checking from these professionals has been widely used due to its prominent and measurable impact [40, 65], this process typically does not involve engaging with the actors spreading misinformation. Instead, the ordinary users' counter-misinformation efforts complement those from professional fact-checkers by directly engaging in countering

Reinforcement Learning-based Counter-Misinformation Response Generation: A Case Study of COVID-19 Vaccine Misinformation

Bing He
Georgia Institute of Technology
Atlanta, Georgia, USA
bhe46@gatech.edu

Mustaque Ahamad
Georgia Institute of Technology
Atlanta, Georgia, USA
mustaq@cc.gatech.edu

Srijan Kumar
Georgia Institute of Technology
Atlanta, Georgia, USA
srijan@gatech.edu

ABSTRACT

The spread of online misinformation threatens public health, democracy, and the broader society. While professional fact-checkers form the first line of defense by fact-checking popular false claims, they do not engage directly in conversations with misinformation spreaders. On the other hand, non-expert ordinary users act as eyes-on-the-ground who proactively counter misinformation – recent research has shown that 96% counter-misinformation responses are made by ordinary users. However, research also found that 2/3 times, these responses are rude and lack evidence. This work seeks to create a counter-misinformation response generation model to empower users to effectively correct misinformation. This objective is challenging due to the absence of datasets containing ground-truth of ideal counter-misinformation responses, and the lack of models that can generate responses backed by communication theories. In this work, we create two novel datasets of misinformation and counter-misinformation response pairs from in-the-wild social media and crowdsourcing from college-educated students. We annotate the collected data to distinguish poor from ideal responses that are factual, polite, and refute misinformation. We propose MisinfoCorrect, a reinforcement learning-based framework that learns to generate counter-misinformation responses for an input misinformation post. The model rewards the generator to increase the politeness, factuality, and refutation attitude while retaining text fluency and relevancy. Quantitative and qualitative evaluation shows that our model outperforms several baselines by generating high-quality counter-responses. This work illustrates the promise of generative text models for social good – here, to help create a safe and reliable information ecosystem. The code and data is accessible on https://github.com/claws-lab/MisinfoCorrect.

CCS CONCEPTS

• **Computing methodologies** → **Natural language generation**; **Reinforcement learning**.

KEYWORDS

misinformation, reinforcement learning, text generation

WWW '23, April 30–May 04, 2023, Austin, TX, USA

ACM ISBN 978-1-4503-9416-1/23/04.
https://doi.org/10.1145/3543507.3583388

ACM Reference Format:
Bing He, Mustaque Ahamad, and Srijan Kumar. 2023. Reinforcement Learning-based Counter-Misinformation Response Generation: A Case Study of COVID-19 Vaccine Misinformation. In *Proceedings of the ACM Web Conference 2023 (WWW '23), April 30–May 04, 2023, Austin, TX, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3543507.3583388

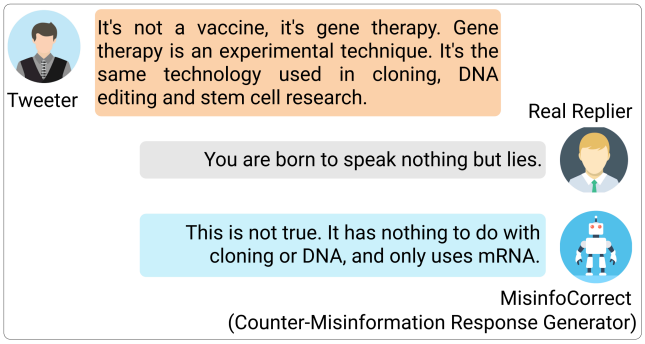

Figure 1: An overview of counter-misinformation response generation task.

1 INTRODUCTION

Online misinformation reduces trust in vaccines and health policies [7, 45, 67], leads to violence and harassment [6, 83], questions democratic processes and elections [79–81], increases polarization [85], and harms well-being [96]. Most people receive information and news from social media [107], which is often "ground-zero" for health misinformation and where misinformation spreads faster and farther than truth [45, 100]. COVID-19 vaccine misinformation, including false claims that the vaccine causes infertility, contains microchips and even changes DNA and genes has fueled vaccine hesitancy, reduced vaccine uptake, and prolonged the pandemic. Besides, misinformation also causes harms to people directly. For example, misinformation that Bill Gates creates vaccines to depopulate people led to distrust and verbal attacks [25]. Thus, it is critical to curb the spread of online misinformation [13, 28, 46, 49, 57, 111, 119]. In this work, we use a broad definition of misinformation which includes falsehoods, inaccuracies, rumors, decontextualized truths, or misleading leaps of logic [114].

Professional fact-checkers and journalists provide objective fact-checks for viral claims and release their determination on their website, which are incredibly useful to create detection models. However, fact-checkers do not actively engage with misinformation