# Unraveling the Potential of Crowds: Insights from a Comparative Study on the Variety, Velocity, Veracity, and Viability of Crowdsourced and Professional Fact-Checking Services

Andy Zhao and Mor Naaman

**Abstract.** The tension between the increasing need for fact-checking and the limited capacity of fact-check providers inspired several crowdsourced approaches to address this challenge. However, little is known about how effectively crowdsourced fact-checking might perform in the real world at a large scale. We fill this gap by evaluating a Taiwanese crowdsourced fact-checking community and two professional fact-checking sites on four dimensions: variety, velocity, veracity, and viability. To this end, we first matched the fact-checking request on the crowdsourced website with professional articles by text similarity. Then we leveraged natural language processing, exploratory data analysis, and manual annotation to evaluate the contributions of these two fact-checking sources along these dimensions. Our results show the different focus these two types of fact-checking approaches have in terms of topic coverage (variety) and demonstrate that while crowdsourced fact-checkers are much faster than professionals (velocity) to answer new requests, these fact-checkers often build on the existing professional knowledge for repeated requests. In addition, our findings indicate that the accuracy of the crowdsourced community (veracity) is comparable to that of the professional sources, and that the crowdsourced fact-checks are perceived by raters to be quite close to professionals in terms of objectivity, clarity, and persuasiveness (viability).

## 1 Introduction

Fact-checking is an important counterstrategy to rampant misinformation online, as it may reduce misconceptions in misleading content or help people correctly evaluate a claim's veracity (Young et al. 2018; Walter et al. 2020). However, fact-checkers, who are often professional journalists (Graves and Amazeen 2019), face a capacity issue dealing with the significant amount of potential misinformation on the internet (Micallef et al. 2022). One proposed way to address this challenge is through crowdsourced fact-checking, where amateur fact-checkers participate in the process of evaluating

*Journal of Online Trust and Safety*, September 2023, page 1 of 24

doi:10.54501/jots.v2i1.118

veracity and provide their feedback as fact-checks (Pennycook and Rand 2019; Allen et al. 2020).

Crowdsourcing services can motivate people to collaborate and create useful knowledge (Sunstein 2006; Kittur and Kraut 2008). The most prominent example is of course Wikipedia, and early studies showed the coverage and the quality of this crowdsourced encyclopedia is comparable to Encyclopedia Britannica (Giles 2005; Sunstein 2006), but that at the same time it introduces different biases (Greenstein and Zhu 2018).

One can consider two models of crowdsourced fact-checking: one in which ordinary people act as participants to make an aggregated judgment (i.e., “wisdom of the crowds”) on the veracity of content (Pennycook and Rand 2019; La Barbera et al. 2020); and another in which dedicated individuals are motivated to create new knowledge contributions in the form of fact-checks (Priedhorsky et al. 2007). We focus on the latter model in this work.

Multiple efforts developed “wisdom of the crowd” approaches for fact-checking. For example, Truthsquad was a 2010 crowdsourced fact-checking experiment, led by fact-checking community Newstrust. The site examined controversial claims and invited evidence and edits from users (Florin 2010b; Pinto et al. 2019). A more recent crowdsourced fact-checking effort is Twitter’s Birdwatch, a community-based system to mobilize users to write and rate notes for suspicious tweets (Coleman 2021). Related studies have suggested that Birdwatch is an effective fact-checking method, though it has some shortcomings such as low consensus and partisan focuses (Pröllochs 2022; Allen, Martel, and Rand 2022).

Cofacts represents the second approach to crowdsourced fact-checking. Rooted in the civil tech culture in Taiwan, Cofacts is a web-based online community dedicated to crowdsourced fact-checking (Haime 2022). It gathers and motivates many enthusiastic volunteers, and regularly responds to different fact-check requests posted by users. In this work, our goal is to explore whether the fact-checking model in Cofacts, which motivates the best rather than finding the average of the crowd, is a promising approach to combat misinformation in the real world, especially compared to professional fact-checking sites.

To undertake this exploration, we need a comprehensive and empirical perspective to evaluate the contributions of fact-checking services. Nieminen and Sankari (2021) developed detailed criteria for choosing and checking processes to evaluate the different aspects of fact-checking content, which mostly reflect on its veracity. Godel et al. (2021) chose to use the labels from professional fact-checkers as the benchmarks and examine how the judgments from a crowd of ordinary people compare to the professionals’ answers. Comparative studies with other professional and crowdsourced services are useful because they may expose the advantages and disadvantages of different models. For example, to compare the quality of Wikipedia with traditional encyclopedias, scholars utilized user or expert reviews, as well as behavioral features like the number of edits, word count, user reputation, and lexical cues (Giles 2005; Wilkinson and Huberman 2007; Hu et al. 2007; Javanmardi and Lopes 2010; Xu and Luo 2011).

In this study, we aim to measure the value of crowdsourced and professional fact-checks along four dimensions:

- Variety: the breadth of fact-check coverage.
- Velocity: the speed of fact-checks.
- Veracity: the reliability and objectivity of fact-checks.
- Viability: the persuasiveness and effectiveness of fact-checks.

This evaluation framework is not normative: it does *not* consider which dimension is more important, nor does it necessarily judge better or worse performance in each dimension.

We applied our analysis framework to evaluate the contributions of Cofacts, in comparison to two well-known professional fact-checking sites in Taiwan, MyGoPen and Taiwan FactCheck Center. We first used Jaro-Winkler (JW) similarity to match fact-checking requests on Cofacts with fact-check articles on the professional sites, which gave us comparable pairs of crowdsourced answers and professional fact-checks. After that, we used three different methods to evaluate the *variety* of topics covered by both sources: topic classification based on user-generated crowdsourced labels, topic clustering based on professional articles, and pairing with user-generated crowdsourced labels. To evaluate *velocity*, we looked at the first request on Cofacts as the demand for a fact-check, and took advantage of timestamps on both sites to find which side responded faster and explore whether professionals had fact-checked a story before the Cofacts request was made. Next, we took professional fact-checks as ground truth and extracted the aggregated response from Cofacts to examine the *veracity* of crowdsourced responses. Last, we asked native speakers to annotate their perceived qualities of fact-checks on three elements ("objectivity," "clarity," and "persuasiveness") in a survey without telling them the true source of fact-checks, to understand the *viability* of the crowdsourced answers.

Our results for the variety dimension suggest that the two types of fact-checking have distinctive focuses: professional fact-checking is more likely to step into fields requiring expertise like medical issues, while crowdsourced fact-checking tends to cover politics or common misinformation in everyday life like fraud messages. The results for velocity show that, as anticipated, crowdsourced fact-checking tends to be faster than professional fact-checking in responding to requests, though professional contributions often provide ready-made answers for crowdsourced fact-checkers to use for recurring fact-check requests. For the veracity dimension, we show that the aggregated claims from multiple crowdsourced answers from Cofacts are almost as reliable as those from professionals: verifiable answers on Cofacts agree with professional fact-checkers 98.8% of the time. Finally, based on the manual annotations of perceived qualities of fact-checks, our analysis for viability shows that crowdsourced fact-checks are almost on par with fact-check articles created by professionals regarding persuasiveness and objectivity measures. Crowdsourced articles even have a small advantage in clarity.

## 2 Background

While scholars may have differing opinions about the effectiveness and benefits of fact-checking, several research contributions have argued that fact-checking is useful to counter some negative implications of misinformation under different cultural contexts (Porter and Wood 2021). A number of studies addressed the evaluation of professional fact-checking quality. Lim (2018) chose to assess the consensus of fact-checking statements and topic coverage and suggested that fact-checkers actually have relatively low topic overlaps and widely variable consensus rates due to different conversion methods. Nieminen and Sankari (2021) designed a list of 24 detailed criteria for fact-check practices and manually examined 858 fact-checks from PolitiFact, concluding that PolitiFact is generally of high quality but has a problem of clearness, as the complex propositions in fact-check claims may confuse users. We borrowed the ideas and logic in the above literature to develop our own evaluation framework.

Indeed, professional fact-checking, limited by nature, faces the significant challenge of keeping pace with enormous levels of misinformation online (Allen et al. 2020),leading scholars to turn to crowdsourced fact-checking as a solution. For example, Florin (2010a) suggested that, based on the Truthsquad experiment, collaboration between professionals and amateurs could deliver reliable fact-checking results. But the credibility of crowdsourced fact-checking still remains a concern, and relevant studies have shown mixed results on this question. La Barbera et al. (2020) suggested that crowds in their experiment exhibit bias in fact-checking, though the aggregated conclusion of crowds is close to experts'. Pennycook and Rand (2019) recruited participants from online platforms and asked them to rate news sources, and suggested that laypeople's judgments about news source quality are very effective if aggregated in a balanced manner, though not as good as professional fact-checkers. Godel et al. (2021) chose to recruit ordinary individuals and professional fact-checkers to evaluate popular news stories. They found that while ordinary users cannot reach the level of professional fact-checkers, machine learning models perform better at identifying false news if trained only on labels from users with a high level of political knowledge. This result suggests that a selective sample of crowdsourced fact-checkers could be helpful in identifying unreliable news.

As Geiger et al. (2011) suggested, crowdsourcing studies can be classified by many characteristics, and contributor group composition and result integration strategy are two important dimensions. However, most studies about crowdsourced fact-checking have used experimental settings and recruited ordinary people as participants to make an aggregated judgment (Pennycook and Rand 2019; La Barbera et al. 2020). Although this is a feasible way to assess the potential of crowds, the results are not the same as the contributions of crowdsourced fact-checking in reality, because different coordination methods may have different outcomes (Kittur and Kraut 2008). In other words, previous literature focused more on a crowdsourcing model that collects information from a crowd of ordinary individuals and aggregates their results as the final judgments, but this approach is only applicable when most people tend to have a correct answer better than a random guess, as Condorcet's jury theorem suggests, which is not always true (Sunstein 2006).

Therefore, our study here focuses on a different crowdsourcing strategy: motivating dedicated individuals to make meaningful contributions, similar to the model used by Wikipedia, as suggested by Godel et al. (2021). The active users in the crowdsourcing community are usually self-selected and have a wide array of motivations (Oreg and Nov 2008). However, the aggregation strategy used in the crowdsourced site we study follows neither the simple average of all opinions nor the Wikipedia model of collaborative editing. Instead, multiple fact-checks can be contributed, then up- or down-voted. Users may read all answers and selectively accept fact-checks with more upvotes.

Hassan et al. (2019) examined such a fact-checking model on Reddit and suggested that comments from ordinary users did provide informative feedback. Amateur fact-checkers played different roles than professional journalists and coordinated with other users to produce effective answers for fact-checking requests. Hassan et al. (2019) suggested that such a crowdsourced fact-checking model, along with help from professionals and automation, has strong potential in the future. Saeed et al. (2022) also compared the crowdsourced fact-checks from Birdwatch with experts on ClaimReview. Their study focused on topic selection, evidence sources, and accuracy, and suggested the differences and advantages that crowdsourced fact-checking presents.

Current literature in fact-checking overwhelmingly focuses on the Western experience, and more specifically the United States. While many Asia-Pacific countries have rich experience battling misinformation by, for example, establishing special governmental agencies or mobilizing civil societies, their practices are largely understudied (Davis, Crowley, and Corcoran 2019; Cha, Gao, and Li 2020). Taiwan offers a very specific context given its approach and challenges in protecting its information ecosystem. To address its challenges of misinformation and ideology clashes, the Taiwanese government collaborated with civil organizations and tech communities to develop digital tools (Cha, Gao, and Li 2020; Chang, Haider, and Ferrara 2021). Our study object in this paper, Cofacts, is one of the most prominent outcomes of government-civil collaboration efforts.

## 3 Method

To compare the contributions of crowdsourced and professional fact-checking, we first collected fact-checks from crowdsourcing and professional sites. We then matched fact-checks about the same requests from both sources to generate comparable pairs for further analysis. This section details our data and matching strategy, as well as a rating task performed to understand readers' perceptions of fact-check articles.

### 3.1 Data

Our context of this study is Taiwan, specifically using a popular crowdsourced fact-checking community called Cofacts. We obtained user-generated request and answer data from Cofacts as our dataset of crowdsourced fact-checking. To construct an equivalent dataset of professional fact-checking, we collected fact-check articles from two popular professional fact-checking sites, MyGoPen and Taiwan FactCheck Center. All text data are in traditional Chinese.

#### 3.1.1 Cofacts

Cofacts is an online fact-checking community founded in 2017. It originated from the Taiwanese decentralized civic tech community *g0v*. In this Quora-like community, all users can either make a request to fact-check a suspicious claim or answer any such requests. In addition, users can second a request, and upvote or downvote an answer.

Cofacts made all data publicly available on GitHub upon request. When we made our data request at the end of July 2021, the site had more than 60,000 fact-checking requests and more than 55,000 fact-checking answers. For most of its operational period, barring an initial launch period, Cofacts' fact-checks (supply) maintained a consistent and relatively steady rate of responses for the number of requests (demand) made on the platform, though the rate decreased towards the end of our data collection period (May 2021), possibly due to the first outbreak of COVID-19 in Taiwan.

According to our data, 1,823 unique Cofacts users answered fact-checking requests at least once. 721 of them answered more than once, and 411 of them answered more than twice. As expected for social contribution platforms, participation is heavily skewed. Around 6% of all users produced 94% of all replies, and 1% of users were responsible for about 90%. Given the voluntary and self-motivated nature of Cofacts, a small number of users may have a significant impact on the community, its contributions, and our analysis. We therefore classified all Cofacts users into three tiers based on their contributions, and provide analysis in this work that considers these tiers separately, where it applies. Tier 1 has only one account. This prominent user has been active since the beginning of this platform up to the end of our data collection and contributed about 48% of all replies on this platform alone. Tier 2 has eight users who have contributed more than 1,000 answers individually (the most active user in this tier has around four thousand answers in our dataset). In total, they contributed more than 32.6% of all Cofacts answers. The rest of the Cofacts users are in Tier 3 of our analysis below.

Cofacts also enables fact-checkers to use different labels to indicate their conclusions. The label "contains misinformation" was used in 44.9% of fact-checking replies, and "contains correct information" was used in 22.2%. "Not fact-checkable" was associated with 19.8% of replies, and "opinionated" with 13.0%.

#### 3.1.2 Taiwanese professional fact-checking sites

We chose two popular professional fact-checking organizations in Taiwan as our sample of professional fact-checkers: MyGoPen and Taiwan FactCheck Center, which publish fact-checking articles on their websites. We collected all public fact-checking articles from the date of creation of these two sites, along with necessary features like dates and labels. We crawled 1,687 articles from MyGoPen and 944 articles from Taiwan FactCheck Center by looping their web pages. The dates of these articles ranged from November 2015 to May 2021.

### 3.2 Matching fact-checking requests and answers

Kazemi, Garimella, Gaffney, et al. (2021) suggested that claim matching between fact-checks of multiple languages is a significant challenge to scaling up global fact-checking. This is also a fundamental challenge for this research, as we need to find comparable pairs between the crowdsourced fact-checks and professional fact-checks on the same topics in order to evaluate variety, velocity, and veracity.

To achieve this goal, we designated the Cofacts request as the "headline" of a fact-check, and the Cofacts responses to this request were designated as the articles under this headline. We then calculated the text similarities between Cofacts requests with the title and summary of professional articles on fact-checking web pages, which returned another article under the same headline.[1] By doing this, we could find pairs of crowdsourced and professional fact-checks that were under the same headline, or that responded to the same content. For instance, a Cofacts question about a hacking virus on a Christmas greeting picture should be paired with Cofacts responses under this request and professional fact-checks about this specific topic, rather than on the COVID-19 virus during Christmas or a hacking virus carried by an email.

To capture the subtle distinctions between seemingly identical Chinese text, we chose Jaro-Winkler (JW) similarity, which measures the edit distance between two strings and is ideal for Chinese characters. To test the validity of JW similarity, we sampled 200 matched article pairs from our dataset (100 positive cases and 100 negative cases with above 0.7 and below 0.6 JW similarity scores, respectively). We asked two native Chinese speakers to annotate the homogeneity in the meaning of two matched articles (i.e., do they talk about the exact same issue). The result indicates that JW had 2 false positives and 37 false negative cases, which means that it had a precision of 0.98 and a recall of 0.73. Since we needed to aim for accurate matches and weigh less on recall in this research, the JW algorithm was a suitable tool for us to distinguish the nuanced differences among Chinese text and find identical fact-checks. We also informally evaluated BERT and realized that distinguishing similar Chinese text is a weakness of this language model.

1. Different users may check the claim in different ways with distinctive languages, or simply cite a web page in Cofacts responses, but a relevant fact-check probably cannot avoid the exact terms in the original content. Therefore, we used crowdsourced requests rather than answers for our calculation to retrieve similar professional articles.

However, the similarity threshold and the time frame to retrieve fact-check candidates may both have great implications on the matching results. Therefore, we tested the JW algorithm with different thresholds (0.6 and 0.7) and several time differences (7, 15, 30, 45, and 60 days) based on our observations and experiences. The results were relatively close, so we chose 0.6 as the similarity threshold and a 45-day difference as our matching window to retrieve more fact-checks. Therefore, we identified 1,222 unique professional fact-checks and 1,496 unique crowdsourcing fact-checks on similar issues and posted within 1.5 months from each other (one professional article may match with multiple crowdsourcing fact-checks). These matched fact-checks were made up of 46.4% of all professional articles and 2.4% of all crowdsourced fact-checks. Since we only used matched pairs for a part of our evaluation, our conclusions are insensitive to the match rates, and these pairs still help us understand the performances of different sources. For example, to evaluate velocity, we evaluated which side was faster to publish a headline from the matched pairs.

## 4 Results

Having set up the data sources and some of the data collection details, we now turn to present an evaluation of the crowdsourced and professional fact-checking based on the four evaluation dimensions. For each, we provide details of how we operationalized the comparison, then present the outcome of the evaluation.

### 4.1 Variety

Our first dimension, variety, represents the topic coverage of fact-checking articles. Our measures for the variety dimension aim to show the differences between the topics covered by the two fact-checking sources. Due to the different volumes of production, we focus on the proportions of topic coverage rather than the topic counts, which are less sensitive to the topic “resolution.”

Knowing the topic distributions from these sources can help us understand not only how the resources are allocated in different fields, but also how crowdsourced and professional fact-checkers engage in the issues of public interest and build trust in their work. We therefore investigated the variety, operationalized as the diversity of topics, as a dimension of interest. We started with the topic distribution analysis not only because it is the best way to demonstrate our matching method, but also because this is the basis for the analysis that follows, which also examines the distinctive performances of two fact-checking providers over different topic fields. We did this analysis in three different ways to increase the robustness of our findings: (1) using user-generated topic labels from crowdsourced fact-checks and supervised learning to predict the topics of professional fact-checks; (2) using BERT embedding and unsupervised learning to cluster professional articles and then predict the topics of crowdsourced fact-checks; and (3) using user-generated labels for a “match and assign” strategy.

#### Topic classification

For the first approach to understanding the topic coverage of fact-checks from both sources, we built on the Cofacts user-generated and -curated labels, assigned to fact-check requests by Cofacts community members. Since the professional fact-checks do not have associated topic labels, we used supervised learning to assign labels to them. To this end, we used crowdsourced fact-checks with topic labels as a training dataset to develop a classification model to infer the topics of professional articles. While this one-sided approach did not allow us to study topics exclusively covered by professional sites, we do not believe such topics were prevalent given our observations and the wide interests of users.

For better classification accuracy, we filtered out topics that had a relatively small amount of cases or had an “unclassified” label. We then balanced our training dataset by oversampling smaller topics and undersampling bigger topics. This process resulted in 23,569 fact-checks with seven topic labels. We then used BERT, a pretrained language model, to compute the embedding features of the text of professional (titles and summaries) and crowdsourced (response bodies) fact-checks (Devlin et al. 2018). This method converts the written text into a form that can be processed mathematically, represented as points in a multidimensional space. With these embeddings, we used a neural network to train our model. Our training sample consisted of 20,000 random fact-checks, and the remaining data were used as our evaluation dataset. The evaluation accuracy on the set of 3,569 article was 0.899 (0.898 precision; 0.903 recall).

The classification results of our model on professional fact-checks are shown in Figure 1 on the following page, along with the topic distributions of crowdsourced fact-checks. A Chi-squared test ($N$=28,048) comparing the article distributions over these topic labels showed no significant differences between the topic distributions of professional fact-checks and crowdsourced fact-checks ($p$=0.99, $\chi^2$ = 0.34, $df$ = 6). However, the high proportion of crowdsourced fact-checks that refer to professional articles to answer a question about fraud messages also indicates that such answers were largely supported by valuable fact-checks from professionals. This phenomenon suggests that while professional fact-checkers did occasionally respond to some requests about fraud messages, recurring needs in this field were usually fulfilled by crowdsourced fact-checkers who helped with the further distribution of those professional fact-checks.

In addition, the high referenced proportion of professional fact-checks in “Health and Food Safety” and “China” indicates the diverse needs of the masses that cannot be satisfied solely by crowdsourced fact-checkers. While not too many crowdsourced fact-checks chose to refer to a professional article on these topics, the contrast between the reference ratios in the two types of fact-checks suggests a reliance of crowdsourced fact-checkers on professional sites to answer some less common but more broad requests.

### Topic clustering

In the second analysis of the variety of fact-checks, we used topic clustering to evaluate the fact-check topic distributions from professional and crowdsourced fact-checks.

In contrast to the first step, where we built a model on crowdsourced fact-checks, here the analysis was driven by the data from professional fact-checkers using K-Means clustering. Simply put, this method groups together professional fact-checks that are similar to each other based on the features of their text. To do that, we again used the embedding values of BERT to represent the fact-check text. These clusters were created by measuring the distance between different texts in the multidimensional embedding space. In this process, we experimented with different numbers for $k$, representing the number of groups we wanted to create. We found that $k$ = 5 led to the best grouping, as it scored highest on the silhouette score (a metric to evaluate the quality of clustering) and showed the best cohesion within groups. This process resulted in five distinct clusters. We then used these clusters as an unsupervised model to predict the cluster labels of the crowdsourced fact-checks by finding, for each crowdsourced fact-check, the cluster

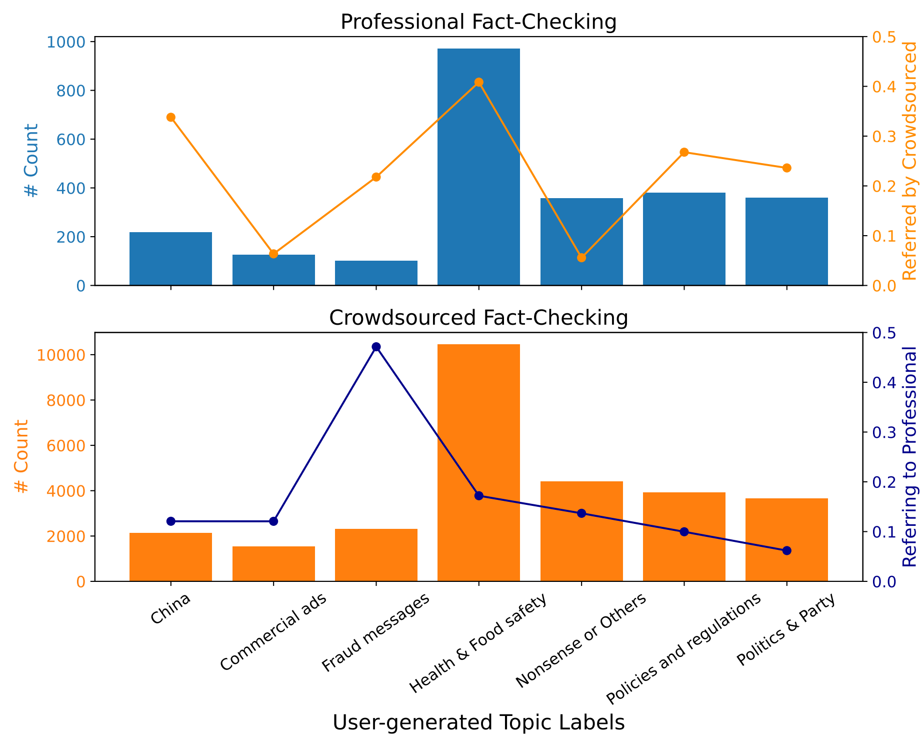


Figure 1: Topic distributions of professional and crowdsourced articles. The top graph represents professional fact-checks; blue bars (left y-axis) show article count number and the orange line shows the number of professional fact-checks that are cited by crowdsourced fact-checks. The bottom graph represents crowdsourced fact-checks: orange bars (left y-axis) show article count number and the blue line (right y-axis) shows the number of crowdsourced fact-checks that cited professional fact-checks.

whose centroid was closest in the embedding space. For presentation, we summarized the fact-check topics by manually examining the contents in each article cluster.

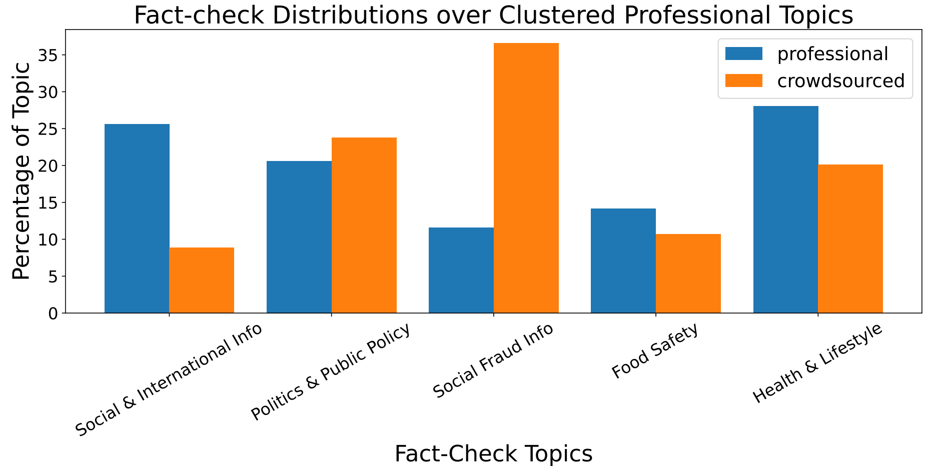


Figure 2: Topic distributions of professional and crowdsourced fact-checking over five topic clusters. The blue bars represent the number of professional fact-checks in each topic, and the orange bars represent the number of crowdsourced fact-checks.

Figure 2 shows the topic distributions of both professional and crowdsourced fact-checks over five topic clusters. A one-way Chi-squared analysis ($N$=63,852) suggests that these topic distributions in five clusters of two fact-checking sources were not significantly different ($p$=0.95, $\chi^2$=0.69, $df$=4). The topics were summarized after we manually examined the fact-checks in clusters, so there are more distinctive topics than in Figure 1 (for example, "Health and Food safety" is separated into two different clusters in Figure 2: "Food Security" and "Health and Lifestyle"). The figure suggests that crowdsourced fact-checkers posted more on "Social Fraud Information"; based on our evaluation, these are topics such as anecdotes, store discounts, missing kids, etc. "Politics and Public Policy" is another topic domain where crowdsourced fact-checkers wrote proportionally more answers. Professional fact-checkers examined more topics like "Social or International information" (for instance, rumors about a Japanese aquarium, counterfeit money, NASA's alien encounters, etc.), "Food Security," and "Health and Lifestyle."

Our supervised and unsupervised learning agree that crowdsourced fact-checkers tend to write more on social fraud messages or policies that are relevant to daily information. However, our two methods show different results about which source may post comparatively more on the topics of health, lifestyle, and food safety, while all our findings confirm that these topics are very popular.

### Match with user-generated labels

Our third technique to examine topic distributions of fact-checks still utilizes user-generated topic labels on Cofacts, but overcomes the lack of labels in the professional dataset by computationally matching professional articles with crowdsourced requests, as explained above in the Method section. We assigned the topic labels in Cofacts requests to corresponding crowdsourcing answers *and* the matched professional articles. This analysis aimed to understand, for all crowdsourced topic labels, which one also had specific stories covered (more or less) by professionals.

After deduplication at the article level to exclude recurring requests and repeat answers, we calculated the topic distributions of crowdsourced and professional fact-checking.

Table 1: Topic distributions in Cofacts and matched professional articles

| Domain | # of Cofacts | # matched with Pro | match rate (%) |
|---|---|---|---|
| Covid-19 | 365 | 37 | 10.14 |
| Technologies and privacy | 327 | 21 | 6.42 |
| Medical issues | 81 | 5 | 6.17 |
| LGBT and AIDS | 286 | 14 | 4.90 |
| Environment protection | 364 | 17 | 4.68 |
| Health and Food security | 11,747 | 418 | 3.56 |
| Agricultural policy | 481 | 15 | 3.12 |
| Policies and regulations | 4,093 | 127 | 3.10 |
| China | 2,358 | 72 | 3.05 |
| Fraud messages | 2,431 | 72 | 2.96 |
| Signing and donating | 570 | 14 | 2.46 |
| Gender issues | 240 | 5 | 2.08 |
| Politics and parties | 4,308 | 77 | 1.79 |
| Commercial ads | 1,560 | 21 | 1.35 |
| Electric and energy | 186 | 2 | 1.08 |

Table 1 shows the results in the topics with at least 150 fact-checks and suggests that both types fact-checked many suspicious stories on issues like health and food security, China, and regulations.

Given the fact that professional fact-checking articles matched 2.4% of the crowd-sourcing dataset with a 0.6 JW threshold, we treated the topic with higher matching rates as the domain where professionals paid more attention to, and vice versa. In other words, professionals checked proportionally more stories in some topics compared to crowdsourced fact-checkers, and these topics were more likely to have a higher match rate than expected. Table 1 suggests that professional fact-checkers tended to focus more on COVID-19, technologies and privacy, environment protections, and other medical issues, and crowdsourced fact-checking were more likely to write articles in response to requests on the topics about fraud messages, energies, and political parties. This result is also consistent with our observations in Figure 2.

Last, we also examined whether the power users (Tier 1 and Tier 2, as explained in the Data section) have similar topic focus, or whether they have the potential to skew the conclusion. We did three Chi-squared analyses on all topic distributions of answers between all three tiers of users, based on all types of Cofacts user labels. All three Chi-squared analyses suggest no significant difference between topic focuses in each tier ($p$=1.0 and $df = 6$ for all). The $\chi^2$ are 0.09 (Tier 1 vs Tier 2, $N$=49,486), 0.11 (Tier 1 vs Tier 3, $N$=40,949), and 0.03 (Tier 2 vs Tier 3, $N$=31,743).

Overall, the three distinct analyses we performed to evaluate the variety of crowdsourced and professional fact-checking show that professional fact-checkers tend to examine the information that requires some knowledge or has bigger implications, for example, medical or health news and international affairs. On the other hand, crowdsourced fact-checkers are proportionally more likely to focus on recurring fraud messages or local political news.

### 4.2 Velocity

Our next dimension, velocity, represents the response speed of fact-checking articles. Since this speed fact-checks matters (Brashier et al. 2021), a faster reaction and earlier fact-check response to potential misinformation could be highly valuable and help limit its spread.

To compare the response speed of the different services, we again took the Cofacts data as a baseline, using the *requests* for fact-checks as "time zero" for global fact-check needs. On Cofacts, we took the time difference between the original request time and the first response to the request as its response time.

For the professional fact-checking sites, we identified the articles that match the Cofacts request, and used the time difference between the first Cofacts request and the corresponding professional article as its response time.

However, relying on Cofacts requests is only an approximation of real-world demand for new fact-checks. Professional fact-check articles could already exist but be unknown to users of Cofacts (if there is a similar request on Cofacts, users may find it with auto-searching during the reporting or notice it in the "similar suspicious message" section). Luckily, the crowdsourcing community itself helps us address this challenge.

We divided our matched fact-check pairs into two parts: professional articles that existed before the request and those written after the request. In the first part, crowdsourced contributors took advantage of fruits planted by professional fact-checkers by responding to requests with a citation to a professional fact-check article. In fact, 454 out of 897 matched fact-check pairs (with 0.6 JW similarity threshold) on Cofacts who had existing answers chose to directly cite a link to MyGoPen or Taiwan FactCheck Center to answer these outdated requests. This result indicates that crowdsourced fact-checkers rely on their professional counterparts to respond to recurring requests to a large extent.

We then analyzed the requests that did not have a ready answer on professional sites. We treated these requests as a "clean slate," assuming that there was no previous fact-check on the topic but that they were also checked by professionals later. Even after excluding those requests in the first part (which is roughly half of all requests on Cofacts), our results indicate a clear advantage for crowdsourcing fact-checking in answering emerging demands. In general, Cofacts was faster in 754 cases out of 879 "clean slate" cases.

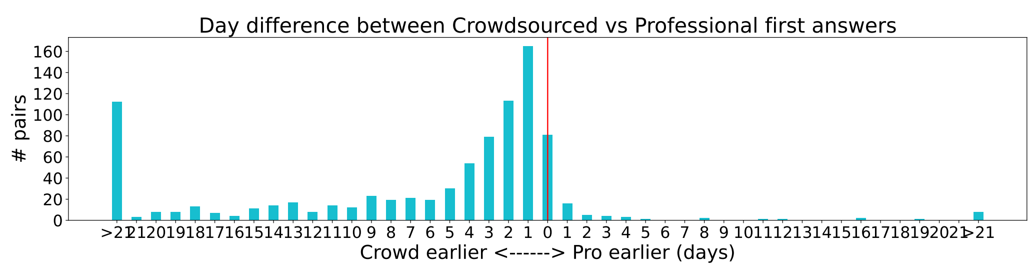


Figure 3: The reaction time differences between professionals and crowdsourcing. The left part of this bar plot represents the cases where crowdsourced fact-checks were faster; the right part indicates where professionals were faster.

Figure 3 shows the day difference distributions between professional and crowdsourced fact-checkers, which suggests that crowdsourced fact-checkers outpaced professionals in velocity by a large extent. Table 2 demonstrates the contributions of crowdsourced and professional sites in rapid and slow cases: crowdsourced fact-checks were earlier than their counterparts in both circumstances. ("Rapid case" means that at least one

Table 2: Velocity comparison between crowdsourced and professional fact-checking

| Faster | Crowdsourcing | Tie | Professional |
|---|---|---|---|
| Rapid cases | 502 | 168 | 25 |
| Slow cases | 116 | 28 | 40 |

site responded to the fact-checking request within 24 hours; “Slow case” means that both sites answered the request after 24 hours; “Tie” means that the professionals and crowdsourcing sites answered the request on the same day.) If we regard rapid cases as easy questions and slow cases as questions that require more effort to explore, our data also imply that professional fact-checkers may have a comparative advantage in requests that may take more time.

In addition, if there was a professional fact-check before the request, the median time for the crowdsourcing community to answer was 17 minutes; if there was never a professional fact-check at all (even after the request), the median time for the crowdsourcing community to respond was about 12 hours longer. However, it is hard to claim any causality here, since it could be that professionals accelerated the fact-checking or that professionals tend to avoid the tricky questions.

The velocity differences also varied among different topics. Table 3 shows the comparisons of matched pairs on different topics, where “pre-answered” means requests that were already answered by professionals. It suggests that the advantage of crowdsourced fact-checking in velocity still holds in different domains, and recurring requests are more likely to happen for topics like COVID-19 and fraud messages.

Furthermore, we also examined the speed differences between different tiers of Cofacts users by checking the percentages of responses within one hour, between one to six hours, between six to twenty-four hours, and above 24 hours, for each tier. As Table 4 suggests, more than 60% of answers from Tier 3 users were at least 24 hours later than the requests, while only around 40% of answers from Tier 1 and Tier 2 users were late by at least 24 hours. In other words, most Cofacts users were indeed slower than those few active users, and the latter group largely improved the response rates on Cofacts. We also compared the crowdsourced fact-checks from only Tier 3 users with the professional fact-checks, excluding the cases where professionals already answered before the request was made on Cofacts. Cofacts users still answered earlier than professional fact-checkers in more than 80% of the cases. In other words, our conclusion that crowdsourced fact-checkers have a clear advantage in velocity over professional fact-checkers still holds, even if we exclude the most active tiers of users, who improve the response speed even more.

### 4.3 Veracity

Our third dimension is veracity, which represents the credibility of fact-checking articles. Fact-checking is expected to be as neutral as possible, but even one of the best crowdsourcing products, Wikipedia, is not as objective as Encyclopaedia Britannica (Greenstein and Zhu 2018). Therefore, a big concern remains over crowdsourcing fact-checking: how reliable such contribution might be compared to professional journalists and fact-checkers.

To analyze the the veracity dimension, we used the *professional* fact-checking as a baseline (or ground truth) for the veracity rating of an article. We measured whether the labels associated with fact-checks by crowdsourced contributions on Cofacts were aligned with

Table 3: Faster response counts in paired case comparisons between crowdsourced and professional fact-checking

| Domain | Crowd | Tie | Pro | Pre-answered |
|---|---|---|---|---|
| Health and Food security | 212 | 14 | 10 | 228 |
| Policies and regulations | 73 | 11 | 4 | 54 |
| Fraud messages | 25 | 7 | 3 | 45 |
| Politics and parties | 32 | 1 | 3 | 43 |
| China | 31 | 5 | 3 | 31 |
| Covid-19 | 23 | 0 | 0 | 20 |
| Commercial ads | 14 | 1 | 1 | 11 |
| Environment protection | 7 | 0 | 2 | 9 |
| Signing and donating | 13 | 0 | 0 | 8 |
| Technologies and privacy | 12 | 1 | 1 | 7 |
| Agricultural policy | 10 | 1 | 0 | 3 |
| LGBT and AIDS | 13 | 0 | 0 | 2 |
| Gender issues | 1 | 0 | 0 | 2 |
| Electric and energy | 2 | 0 | 0 | 0 |

Table 4: Cofacts Users Fact-check Response Time from Different Tiers

| Response Time Ratio | Tier 1 User | Tier 2 Users | Tier 3 Users |
|---|---|---|---|
| <=1 Hour | 0.185 | 0.121 | 0.075 |
| 1–6 Hours | 0.222 | 0.235 | 0.137 |
| 6–24 Hours | 0.187 | 0.255 | 0.177 |
| 24 Hours+ | 0.406 | 0.390 | 0.610 |
| *Total Number* | *29251* | *19558* | *11355* |

the labels assigned to the fact-check article of the same issue by professionals. Only absolute and clear labels on professional fact-checks were taken (in a scale of True and False), which were consistent among the two professional sites and which we turned into a numeric scale. For the crowdsourced fact-checking, we selected the majority opinion of all binary True/False labels under each Cofacts request, which is more natural from a user's perspective.

The matching method explained above identified 663 question-answer pairs from crowdsourced and professional fact-checkers. Out of 643 unique Cofacts requests in matched question-answer pairs, roughly 61% of them received only one answer and 28% of them received two answers. Most of the time, fact-check requests received unanimous answers: 546 requests received all False labels, and 69 requests received all True labels.

After our initial analysis of label alignment, we manually examined the cases where crowdsourced labels were inconsistent with professional labels to validate our results. We found that some disagreements were Cofacts requests asking to corroborate professional fact-checks, which we refer to as "double-check" cases. In other words, professional fact-checkers debunked a rumor (giving a "False" label); and a Cofacts request asked to verify this fact-checking article; then the crowdsourcing users endorsed the conclusion of this professional article (giving a "True" label). Our JW algorithm matched these fact-checks because there are identical texts in Cofacts requests and professional fact-checks.

Therefore, we excluded these "double-check" answers and the cases where a professional article existed before a fact-check request was posted on Cofacts, because crowdsourced fact-checkers can simply copy and paste answers from professional sites. There was only one disagreement in the cases that were already checked by professionals, but it turned out to be a mismatch due to messy characters. For all the cases that were never checked by professionals and that crowdsourced fact-checkers examined independently in advance, the veracity trend remained unchanged in our final result, as shown in the confusion matrix in Table 5. There were only a few disagreements between crowdsourced fact-checking and professional fact-checking.

Table 5: Confusion matrix between professional and crowdsourced fact-check labels after data filtering

| | | Pro | |
|---|---|---|---|
| | | True | False |
| Crowd | True | 12 | 4 |
| | False | 0 | 305 |

The crowdsourced data we evaluated in this section included 222 answers from the Tier 1 user, 75 answers from Tier 2 users, and 66 answers from Tier 3 users (there could be multiple answers in one single fact-check case). Clearly, veracity levels were high throughout. However, there is a concern about amateur contributions in the context of fact-checking: while their fact-checks can have high veracity in most cases, they may be less reliable when evaluating critical issues. We identified four valid cases where professionals and Cofacts users disagreed with each other. The first case was about a church. While the professional fact-checker on MyGoPen and a Tier 3 user on Cofacts had similar answers and references, they used different labels. The second case was about a haze weather warning, answered by the only Tier 1 user and two Tier 2 users. Both sides were correct about this issue, but the timeliness of their answers resulted in different labels. The third case was about the mask policy at voting stations. The two sides—a Tier 3 user on the crowdsourced side—disagreed on whether masks were mandatory. The last case was about a new medical technology. The Cofacts Tier 1 user confirmed the existence of this technology on the market, while MyGoPen confirmed with a hospital that it didn't have such technology. Even in these disagreement cases, the differences between professional and crowdsourced fact-checkers are still not very straightforward most of the time.

In general, while crowdsourced fact-checkers may occasionally disagree with professional fact-checkers, crowdsourced fact-checking still holds a similar level of veracity as professional fact-checkering most of the time.

## 4.4 Viability

The last dimension we consider is viability. This represents how likely it is that the input from fact-checks would be seen as useful by readers. As such, we aimed to evaluate how fact-checks from both crowdsourced and professional sources may be viewed by readers. To do this, we employed evaluators who rated a collection of articles based on various factors. These factors and the entire process are explained in more detail in Section 4.4.1.

### 4.4.1 Annotating perceived quality

Fact-checks are read and understood by humans. As part of our evaluation, we used raters to help us estimate the likelihood of the contributions of fact-checkers being perceived by readers along different measures of quality.

To understand the perceived quality of fact-checks, we randomly sampled hundreds of pairs of crowdsourced and professional fact-check articles on the same false stories, which we manually examined. We selected 40 article pairs (one crowdsourced fact-check and one professional fact-check on the same topic) for this procedure such that the pairs represented balanced topics (15 for medical and health information, 15 for domestic stories, and 10 for international stories); both articles in each pair were original (not just a reference to another site).

We recruited seven native Taiwanese graduate students as raters to read and evaluate these fact-checks. Participants were randomly exposed to either a crowdsourced or professional fact-check article from each of the 40 pairs (without knowing its source). Only text and images in the fact-checks were presented to participants, to avoid the influence of website designs and other factors on the perceptions of fact-checks. Participants were asked to read fact-checks carefully and annotate how they perceived the qualities of each article in three measures, on a scale from 1 to 5: persuasiveness, clarity, and objectivity. As each rater read 40 articles, we obtained 280 responses for each fact-check pair and therefore 140 responses for each (crowdsourced or professional) fact-check article.

### 4.4.2 Measurement results

The first measure is "objectivity." With this measure, our intent was to understand whether fact-checks from both sources are perceived as objective or neutral from a reader's perspective. Our results ($N$=280) show that raters ranked the professional fact-check articles as somewhat more objective. The mean objective rating for professional fact-check articles was 4.16 (SD=0.94), compared to 3.79 (SD=1.16) for the crowdsourced fact-check articles. The difference between the two sets of articles was significant ($p < .01$), though the effect size was relatively small (0.34). The t-statistic was 2.87 and the degree of freedom was 278.

Our second dimension, "clarity," aimed to assess whether a fact-check expresses reasoning and outcome in a simple and understandable way. Our results ($N$=280) show that, in this case, raters ranked the crowdsourced fact-checks (mean=4.24, SD=0.86) as more clear and comprehensive than the professional fact-checks (mean=4.01, SD=1.03). The difference between the two sets of articles was significant ($p < .05$), though the effect size was quite small (0.25). The t-statistic was 2.07 and the degree of freedom was 278.

Finally, our third measure, "persuasiveness," aimed to capture whether readers might find a fact-check strong enough to convince them. Our results ($N$=280) show that raters ranked the professional fact-check articles as somewhat more persuasive. The mean rating for professional fact-check articles was 4.14 (SD=0.96), compared to 3.83 (SD=1.01) for the crowdsourced fact-check articles' persuasiveness. The difference between the two sets of articles was significant ($p < .01$), though the effect size was relatively small (0.32). The t-statistic was 2.65 and the degree of freedom was 278.

In summary, these findings indicate that professional fact-check articles were found to be more persuasive and more objective than crowdsourced fact-checks by our raters, but that crowdsourced fact-checking articles received higher ratings for their clarity. At the same time, the differences between the two sets of articles, while significant, were not substantial. Moreover, all the measures received rating averages of roughly 4.0 for both sets of articles, indicating that fact-checks from both sources were generally perceived as objective, clear, and persuasive.

## 5 Discussion

Our analysis of Cofacts, a crowdsourced fact-checking service, highlights its complementary role alongside professional fact-checkers. Cofacts addresses local and daily affairs, mitigating everyday misinformation effects, while benefiting from the professional fact-checks' global cross-language perspective. It responds more rapidly to fact-checking needs and provides almost as reliable and effective information as professionals.

The variety analysis showed that Cofacts can cover most topics in professional fact-checks and even provide answers to many issues that professionals may skip, though some questions remain about how to interpret the difference in coverage. In particular, we cannot provide any normative judgment on the topic preferences of crowdsourced and professional fact-checkers. The professionals provide a national service and invest more resources into high-visibility issues like COVID-19. Comparatively, crowdsourced fact-checking is more grassroots and driven by crowd-based requests, therefore putting more resources into local and daily affairs and common fraud messages. The breadth of response to community needs on varied topics is a critical offering, as it helps mitigate the effects of "everyday misinformation" users encounter (Lu et al. 2020; Wahlheim, Alexander, and Peske 2020), which may equip readers with cognitive defenses to ward off potential harms from misinformation and reduce the dissemination of suspicious stories (Pennycook et al. 2021; Ecker et al. 2022).

At the same time, the responses from crowdsourced fact-checkers also build on the support of abundant fact-checks on professional sites. Roughly half of the requests responded to on Cofacts could be answered by simply referring to existing fact-checks. Professional fact-checking certainly does not have the capacity to actively respond to an overwhelming number of (often recurring) requests. Under this circumstance, professional fact-checking becomes a manufacturer of knowledge, and crowdsourced fact-checkers play the role of distributors, connecting requests and answers in the information market. Standing on the shoulders of professionals gives crowdsourced fact-checkers critical support in directly countering misinformation, more comprehensively and more quickly.

We note that citing professional fact-checks could also help the crowdsourcing community bring in the perspective of global cross-language fact-checking. We observed in the Cofacts data that crowdsourced fact-checkers occasionally also refer to English professional fact-checking sites like Snopes, sometimes with translated summaries. This contribution is unique and important because manual claim-matching, though not scaled, can largely help with the knowledge dispersion in a cross-language way and counter the misinformation that originates from other countries or is debunked by other fact-checkers (Kazemi, Garimella, Gaffney, et al. 2021). Because sometimes only local fact-checkers have the ability and knowledge to check a story (Ribeiro et al. 2021), reusing fact-checks in other languages can further reduce workloads and increase fact-checking capacity.

The velocity findings follow a similar theme: when crowdsourced fact-checks do not build on the earlier contributions of professionals, we find that they still respond more rapidly to fact-checking needs than professional fact-checkers. Our result holds for both rapid responses (which normally take a few hours) or slow responses (which normally take more than 24 hours), and the advantage of crowdsourced fact-checking is usually as substantial as several days. The result is not affected by the topic. Since misinformation usually spreads quickly on social media (Vosoughi, Roy, and Aral 2018), a faster fact-check response is necessary to avoid greater damage.

Crowdsourcing power could also help with identifying potential misinformation, given its distinctive variety and its advantage in velocity. Messaging platforms like Line and WhatsApp even allow users to report suspicious messages to third-party fact-checkers like Cofacts or to platform fact-checkers (Kazemi, Garimella, Shahi, et al. 2021), which may improve the efficiency of both crowdsourced and professional fact-checking.

Naturally, the veracity of crowdsourced fact-checking is one of the most important concerns. Our data suggest that, taking professional articles as ground truth (of course, itself a challenging proposition), crowdsourced fact-checking can provide answers almost as reliably as professionals. Our veracity findings are also consistent with the conclusion about crowdsourced contributors with higher political knowledge (Godel et al. 2021): mobilizing a more specialized and savvy sample of the population can be a great help for fact-checking. This is not surprising, because the self-selective crowdsourcing model in Cofacts can motivate users with more experience and knowledge to engage more frequently and provide more reliable answers than average people in experiment settings (Pennycook and Rand 2019; Godel et al. 2021; Kaufman, Haupt, and Dow 2022). However, as we further discuss below, there could be long-term challenges in engaging these types of individuals and preventing potential bias and manipulation in the future.

Meanwhile, our viability findings suggest that crowdsourced fact-checking articles are perceived as nearly as persuasive and objective as professional fact-checking, and even perform slightly better on a clarity measure. Those differences were all small in terms of effect size, suggesting that perhaps there is no substantial difference in those qualities between the two types of fact-checking from the readers' perspectives. On the other hand, the significant difference in rates may also imply that the language style, as a medium of fact-checking, could make a difference in convincing readers of its viability (Nieminen and Rapeli 2019). Professional fact-checks are usually longer and contain detailed domain knowledge, which signals their expertise and objectivity but also creates a barrier for various readers to understand. However, simple language can better facilitate the corrections of misbelief (Ecker et al. 2022), and the contribution of the crowdsourced community could help make professional content more accessible to readers. In our annotation task, we excluded all crowdsourced answers that solely cited professional fact-checks. But in practice, this kind of paraphrase or a summary of a professional article by crowdsourced contributors may provide this desired improved accessibility.

Furthermore, the contribution of fact-checking is highly associated with cultural background and social context (Sultana and Fussell 2021). It is noteworthy that the success of Cofacts is associated with its strong foundation in the vibrant Taiwanese civic tech culture and the rich culture and history of vigorous crowdsourcing activism in Taiwan (Ho 2012; Hsiao and Kuan 2016; Lee 2020). Additionally, Taiwan's unique geopolitical context, particularly the ongoing threat of Chinese misinformation, has heightened public awareness and motivated the population to actively engage in combating misinformation (Chang, Haider, and Ferrara 2021; Haime 2022). It is possible that this active engagement, in turn, strengthens the crowdsourced fact-checking efforts, making platforms like Cofacts more robust and reliable in their mission to counter the spread of false information.

Our study also has two practical limitations due to the methods we used in our approach. We relied on the textual matching method to identify the comparable pairs of crowdsourced and professional fact-checks. The potential biases of this approach may impact the results, because the writing habits of individuals from both crowdsourced and professional sides may result in under- or over-identification of paired fact-check cases. Another practical limitation is our rater task, where we asked raters to provide an evaluation of fact-check articles. Since we lack good tools to reach out to non-Western citizens, our raters were recruited from a student population at a prestigious university, and thus do not provide a good sample of the average Taiwanese users.

Our work highlights the value of crowdsourcing communities like Cofacts, which can mobilize dedicated individuals online to counter misinformation on social media along with professional fact-checkers. Our approach assumes that fact-checking is a positive societal contribution, and further assumes that individuals undertake it with a commitment to provide accurate information to the best of their knowledge. However, the concern still remains that a crowdsourced system could be abused by malicious users, as can be observed in many other systems and is also acknowledged by Cofacts (Davis, Crowley, and Corcoran 2019). In the extreme, our work here can inform and motivate such users, although we believe the risk of that is low. Instead, we hope this work can encourage support for crowdsourcing services and highlight the need to protect them, e.g., by preventing inauthentic behaviors or information pollution campaigns (Shachaf and Hara 2010; Rawat et al. 2019).

## 6 Conclusion

Our study provides promising evidence that crowdsourced fact-checking, exemplified by Cofacts, offers valuable and high-quality contributions in combating misinformation. Complementing professional fact-checking services, crowdsourced fact-checking offers distinct but high-quality contributions across multiple dimensions that are comparable to professional fact-checking efforts. Our findings provide a hopeful indication that community-based crowdsourced approaches could offer important support to counter online misinformation, thereby helping to advance a society that is less vulnerable to present and future challenges.

## References


Allen, Jennifer, Antonio A. Arechar, David G. Rand, and Gordon Pennycook. 2020. "Crowdsourced Fact-Checking: A Scalable Way to Fight Misinformation on Social Media." *Collective Intelligence,* https://doi.org/10.48550/arXiv.2208.09214.

Allen, Jennifer, Cameron Martel, and David G. Rand. 2022. "Birds of a feather don't fact-check each other: Partisanship and the evaluation of news in Twitter's Birdwatch crowdsourced fact-checking program." In *CHI Conference on Human Factors in Computing Systems,* 1–19. https://doi.org/10.1145/3491102.3502040.

Brashier, Nadia M, Gordon Pennycook, Adam J. Berinsky, and David G. Rand. 2021. "Timing matters when correcting fake news." *Proceedings of the National Academy of Sciences* 118 (5). https://doi.org/10.1073/pnas.202004311.

Cha, Meeyoung, Wei Gao, and Cheng-Te Li. 2020. "Detecting fake news in social media: an Asia-Pacific perspective." *Communications of the ACM* 63 (4): 68–71. https://doi.org/10.1145/3378422.

Chang, Ho-Chun Herbert, Samar Haider, and Emilio Ferrara. 2021. "Digital civic participation and misinformation during the 2020 Taiwanese presidential election." *Media and Communication* 9 (1): 144–57. https://doi.org/10.17645/mac.v9i1.3405.

Coleman, K. 2021. *Introducing Birdwatch, a community-based approach to misinformation.* Twitter, January. https://blog.twitter.com/en_us/topics/product/2021/introducing-birdwatch-a-community-based-approach-to-misinformation.

Davis, Raina, Bo Julie Crowley, and Casey Corcoran. 2019. "Civil society: A key player in the global fight against misinformation." *Kennedy School Review* 19:171–73. https://ksr.hkspublications.org/2020/02/07/civil-society-a-key-player-in-the-global-fight-against-misinformation/.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. "BERT: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805,* https://doi.org/10.48550/arXiv.1810.04805.

Ecker, Ullrich, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa Fazio, Nadia Brashier, Panayiota Kendeou, Emily Vraga, and Michelle Amazeen. 2022. "The psychological drivers of misinformation belief and its resistance to correction." *Nature Review Psychology* 1:13–29. https://doi.org/10.1038/s44159-021-00006-y.

Florin, F. 2010a. *Crowdsourced Fact-Checking? What We Learned from Truthsquad.* MediaShift, November. http://mediashift.org/2010/11/crowdsourced-fact-checking-what-we-learned-from-truthsquad320/.

———. 2010b. *Introducing Truthsquad.* NewsTrust, August. https://blog.newstrust.net/2010/08/truthsquad.html.

Geiger, David, Stefan Seedorf, Thimo Schulze, Robert C. Nickerson, and Martin Schader. 2011. "Managing the crowd: towards a taxonomy of crowdsourcing processes." *AMCIS 2011 Proceedings — All Submissions* 430. https://aisel.aisnet.org/amcis2011_submissions/430.

Giles, Jim. 2005. "Internet encyclopaedias go head to head." *Nature* 438:900–901. https://doi.org/10.1038/438900a.

Godel, William, Zeve Sanderson, Kevin Aslett, Jonathan Nagler, Richard Bonneau, Nathaniel Persily, and Joshua Tucker. 2021. “Moderating with the Mob: Evaluating the Efficacy of Real-Time Crowdsourced Fact-Checking.” *Journal of Online Trust and Safety* 1, no. 1 (October). https://doi.org/10.54501/jots.v1i1.15.

Graves, Lucas, and Michelle A. Amazeen. 2019. “Fact-checking as idea and practice in journalism.” In *Oxford Research Encyclopedia of Communication,* edited by Oxford University Press. https://doi.org/10.1093/acrefore/9780190228613.013.808.

Greenstein, Shane, and Feng Zhu. 2018. “Do experts or crowd-based models produce more bias? Evidence from Encyclopædia Britannica and Wikipedia.” *Mis Quarterly,* https://doi.org/10.25300/MISQ/2018/14084.

Haime, Jordyn. 2022. “Taiwan’s amateur fact-checkers wage war on fake news from China.” *Al Jazeera* (September). https://www.aljazeera.com/economy/2022/9/19/taiwan.

Hassan, Naeemul, Mohammad Yousuf, Md Mahfuzul Haque, Javier A. Suarez Rivas, and Md Khadimul Islam. 2019. “Examining the roles of automation, crowds and professionals towards sustainable fact-checking.” In *Companion Proceedings of The 2019 World Wide Web Conference,* 1001–6. https://doi.org/10.1145/3308560.3316734.

Ho, Ming-sho. 2012. “Sponsoring civil society: State and community movement in Taiwan.” *Sociological Inquiry* 82 (3): 404–23. https://doi.org/10.1111/j.1475-682X.2012.00413.x.

Hsiao, Hsin-huang Michael, and Yu-Yuan Kuan. 2016. “The development of civil society organizations in post-authoritarian Taiwan (1988–2014).” In *Routledge Handbook of Contemporary Taiwan,* edited by Gunter Schubert, 253–67. Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315769523-21/development-civil-society-organizations-post-authoritarian-taiwan-1988%E2%80%932014-hsin-huang-michael-hsiao-yu-yuan-kuan.

Hu, Meiqun, Ee-Peng Lim, Aixin Sun, Hady Wirawan Lauw, and Ba-Quy Vuong. 2007. “Measuring article quality in wikipedia: models and evaluation.” In *Proceedings of the sixteenth ACM Conference on Conference on Information and Knowledge Management,* 243–52. https://doi.org/10.1145/1321440.1321476.

Javanmardi, Sara, and Cristina Lopes. 2010. “Statistical measure of quality in Wikipedia.” In *Proceedings of the First Workshop on Social Media Analytics,* 132–38. https://doi.org/10.1145/1964858.1964876.

Kaufman, Robert A, Michael Robert Haupt, and Steven P. Dow. 2022. “Who’s in the Crowd Matters: Cognitive Factors and Beliefs Predict Misinformation Assessment Accuracy.” *Proceedings of the ACM on Human-Computer Interaction* 6 (CSCW2): 1–18. https://doi.org/10.1145/3555611.

Kazemi, Ashkan, Kiran Garimella, Devin Gaffney, and Scott A. Hale. 2021. “Claim Matching Beyond English to Scale Global Fact-Checking.” *arXiv preprint arXiv:2106.00853,* https://doi.org/10.48550/arXiv.2106.00853.

Kazemi, Ashkan, Kiran Garimella, Gautam Kishore Shahi, Devin Gaffney, and Scott A. Hale. 2021. “Tiplines to Combat Misinformation on Encrypted Platforms: A Case Study of the 2019 Indian Election on WhatsApp.” *arXiv preprint arXiv:2106.04726,* https://doi.org/10.48550/arXiv.2106.04726.

Kittur, Aniket, and Robert E. Kraut. 2008. “Harnessing the wisdom of crowds in wikipedia: quality through coordination.” In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work,* 37–46. https://doi.org/10.1145/1460563.1460572.

La Barbera, David, Kevin Roitero, Gianluca Demartini, Stefano Mizzaro, and Damiano Spina. 2020. “Crowdsourcing Truthfulness: The Impact of Judgment Scale and Assessor Bias.” *Advances in Information Retrieval* 12036:207. https://doi.org/10.1007/978-3-030-45442-5_26.

Lee, Mei-chun. 2020. “Free the Data from the Birdcage: Opening Up Data and Crowdsourcing Activism in Taiwan.” *PoLAR: Political and Legal Anthropology Review* 43 (2): 247–61. https://doi.org/10.1111/plar.12371.

Lim, Chloe. 2018. “Checking how fact-checkers check.” *Research & Politics* 5 (3): 2053168018786848. https://doi.org/10.1177/2053168018786848.

Lu, Zhicong, Yue Jiang, Cheng Lu, Mor Naaman, and Daniel Wigdor. 2020. “The government’s dividend: complex perceptions of social media misinformation in China.” In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems,* 1–12. https://doi.org/10.1145/3313831.3376612.

Micallef, Nicholas, Vivienne Armacost, Nasir Memon, and Sameer Patil. 2022. “True or False: Studying the Work Practices of Professional Fact-Checkers.” *Proceedings of the ACM on Human-Computer Interaction* 6 (CSCW1): 1–44. https://doi.org/10.1145/3512974.

Nieminen, Sakari, and Lauri Rapeli. 2019. “Fighting misperceptions and doubting journalists’ objectivity: A review of fact-checking literature.” *Political Studies Review* 17 (3): 296–309. https://doi.org/10.1177/1478929918786852.

Nieminen, Sakari, and Valtteri Sankari. 2021. “Checking PolitiFact’s Fact-Checks.” *Journalism Studies* 22 (3): 358–78. https://doi.org/10.1080/1461670X.2021.1873818.

Oreg, Shaul, and Oded Nov. 2008. “Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values.” *Computers in Human Behavior* 24 (5): 2055–73. https://doi.org/10.1016/j.chb.2007.09.007.

Pennycook, Gordon, Ziv Epstein, Mohsen Mosleh, Antonio A Arechar, Dean Eckles, and David G. Rand. 2021. “Shifting attention to accuracy can reduce misinformation online.” *Nature* 592 (7855): 590–95. https://doi.org/10.1038/s41586-021-03344-2.

Pennycook, Gordon, and David G. Rand. 2019. “Fighting misinformation on social media using crowdsourced judgments of news source quality.” *Proceedings of the National Academy of Sciences* 116 (7): 2521–26. https://doi.org/10.1073/pnas.1806781116.

Pinto, Marcos Rodrigues, Yuri Oliveira de Lima, Carlos Eduardo Barbosa, and Jano Moreira de Souza. 2019. “Towards fact-checking through crowdsourcing.” In *2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD),* 494–99. IEEE. 10.1109/CSCWD.2019.8791903.

Porter, Ethan, and Thomas J. Wood. 2021. “The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom.” *Proceedings of the National Academy of Sciences* 118 (37). https://doi.org/10.1073/pnas.2104235118.

Priedhorsky, Reid, Jilin Chen, Shyong (Tony) K. Lam, Katherine Panciera, Loren Terveen, and John Riedl. 2007. "Creating, destroying, and restoring value in Wikipedia." In *Proceedings of the 2007 international ACM conference on Supporting Group Work,* 259–68. https://doi.org/10.1145/1316624.1316663.

Pröllochs, Nicolas. 2022. "Community-based fact-checking on Twitter's Birdwatch platform." In *Proceedings of the International AAAI Conference on Web and Social Media,* 16:794–805. https://doi.org/10.48550/arXiv.2104.07175.

Rawat, Charu, Arnab Sarkar, Sameer Singh, Rafael Alvarado, and Lane Rasberry. 2019. "Automatic Detection of Online Abuse and Analysis of Problematic Users in Wikipedia." In *2019 Systems and Information Engineering Design Symposium (SIEDS),* 1–6. IEEE. https://doi.org/10.1109/SIEDS.2019.8735592.

Ribeiro, Manoel Horta, Savvas Zannettou, Oana Goga, Fabrício Benevenuto, and Robert West. 2021. "Can online attention signals help fact-checkers fact-check?" *arXiv preprint arXiv:2109.09322,* https://doi.org/10.48550/arXiv.2109.09322.

Saeed, Mohammed, Nicolas Traub, Maelle Nicolas, Gianluca Demartini, and Paolo Papotti. 2022. "Crowdsourced Fact-Checking at Twitter: How Does the Crowd Compare With Experts?" In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management,* 1736–46. https://doi.org/10.48550/arXiv.2208.09214.

Shachaf, Pnina, and Noriko Hara. 2010. "Beyond vandalism: Wikipedia trolls." *Journal of Information Science* 36 (3): 357–70. https://doi.org/10.1177/0165551510365390.

Sultana, Sharifa, and Susan R. Fussell. 2021. "Dissemination, Situated Fact-checking, and Social Effects of Misinformation among Rural Bangladeshi Villagers During the COVID-19 Pandemic." *Proceedings of the ACM on Human-Computer Interaction* 5 (CSCW2): 1–34. https://doi.org/10.1145/3479580.

Sunstein, Cass R. 2006. *Infotopia: How Many Minds Produce Knowledge.* Oxford University Press. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=924249.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The spread of true and false news online." *Science* 359 (6380): 1146–51. https://doi.org/10.1126/science.aap9559.

Wahlheim, Christopher N., Timothy R. Alexander, and Carson D. Peske. 2020. "Reminders of everyday misinformation statements can enhance memory for and beliefs in corrections of those statements in the short term." *Psychological Science* 31 (10): 1325–39. https://doi.org/10.1177/0956797620952797.

Walter, Nathan, Jonathan Cohen, R. Lance Holbert, and Yasmin Morag. 2020. "Fact-checking: A meta-analysis of what works and for whom." *Political Communication* 37 (3): 350–75. https://doi.org/10.1080/10584609.2019.1668894.

Wilkinson, Dennis M., and Bernardo A. Huberman. 2007. "Cooperation and quality in wikipedia." In *Proceedings of the 2007 International Symposium on Wikis,* 157–64. https://doi.org/10.1145/1296951.1296968.

Xu, Yanxiang, and Tiejian Luo. 2011. "Measuring article quality in Wikipedia: Lexical clue model." In *2011 3rd Symposium on Web Society,* 141–46. IEEE. https://doi.org/10.1109/SWS.2011.6101286.

Young, Dannagal G., Kathleen Hall Jamieson, Shannon Poulsen, and Abigail Goldring. 2018. "Fact-checking effectiveness as a function of format and tone: Evaluating FactCheck.org and FlackCheck.org." *Journalism & Mass Communication Quarterly* 95 (1): 49–75. https://doi.org/10.1177/1077699017710453.

## Authors

**Andy Zhao** is a PhD candidate in information science at Cornell Tech (Cornell University).

(dandy@infosci.cornell.edu)

**Mor Naaman** Mor Naaman is a Professor at the Jacobs Technion-Cornell Institute at Cornell Tech and in the Information Science Department at Cornell University.

(mor.naaman@cornell.edu)

## Acknowledgements

This material is based upon work partially supported by the National Science Foundation under grant CNS-2120651. We thank Yiling Sun and multiple anonymous Taiwanese annotators for their assistance in this work. We appreciate the reviewers for their helpful feedback. We want to offer our sincere thanks to all those who ever discussed this project with us via email or in person for their insightful comments. Last, we want to express our greatest gratitude to the Cofacts/g0v community and other Taiwanese fact-checkers. This work cannot be done without your devotion to examining suspicious claims, compiling a comprehensive dataset, and building this amazing civic tech community.

## Data availability statement

Replication files are available at: https://osf.io/6rhn8

## Ethical standards

This work is approved by the Institutional Review Board of Cornell University. The IRB protocol number is IRB0010650.

## Keywords

Crowdsourced fact-checking; professional fact-checking; civic tech; computer-mediated communication