

## **Fake News: Just How Reliable Is That?**

Elliot Frank, Xinyun Zhang and Rebecca Gale

### **Abstract:**

*Our project aims to answer the research question: do people who share fake news on Twitter do so knowing it is fake? Our twitter bot offers a framework for this task by first searching for tweets that include a link to a fake news site, and then sending a Twitter reply to the tweet that asks how reliable the user felt the news source they shared was. Our project ran across two difficulties: being blacklisted by Twitter, as well as a low response rate to our tweets. Among the few responses we received, however, several users felt strongly that the fake news site was reliable, indicating that those who read the sites may be distrustful of traditional news outlets. Going forward, we would need to better space the Twitter API to reach more people, and curtail the number of sites tracked to gain better information.*

### **Introduction:**

In the wake of the 2016 presidential election, the concept of “fake news” became widely discussed, debated, and misused in American political discourse. The concept originally referred to online writing that was fantastical but maintained some veneer of truth, that was written in order to be shared widely and therefore generate advertising revenue. Our working definition of fake news is slightly broader, and we are working on the level of entire websites. A website is a “fake news” source if they have shared completely fictional articles and lack basic editorial controls that one would expect in a legitimate newsroom.

Our project attempted to engage with Twitter users who shared articles from the sites we defined as “fake news sites”. Our twitter bot found around 31,000 tweets that shared fake news in the ten-day span that Twitter allows a user to search. Because of the prevalence of such users, we attempted to survey them in an automated way. Here what we call a survey was a single tweet sent as a reply to the tweet where the user shared a fake news link; the tweet asked how reliable they believed the site they were sharing was. The idea was that we could then use the huge volume of data we collected to further gain insight into the attitudes of people who share fake news. We wanted to answer the question: do people who share fake news do so knowing it is fake?

We did not succeed in collecting a huge volume of data, however, or much at all. Our bot was only able to send out ~360 tweets before it was blacklisted by Twitter for violating its terms of service. In addition, we only received 3 responses to our tweets. We attempted to create a second bot, and modified its appearance and messaging to see if it would increase our response rate. It seemingly did get more responses, but after 34 tweets it was also blacklisted. Later in our analysis we discuss these two roadblocks, being blacklisted from Twitter and our low response rate, hypothesize why both of these two difficulties occurred and offer suggestions to researchers to avoid these difficulties. The experience points to the difficulties of trying to

automate certain aspects of journalism, in this case the interview. Trying to reach out to strangers on Twitter is difficult because Twitter sets restraints, and the impersonal nature of the task means there is a low incentive to respond.

### **Methodology:**

Our methodology consisted of three main steps: collection, processing, and replying. As stated before, the input to our script was a list of “fake news sites”. Our list was based on a list of hoax websites compiled by the website Fake News Watch, but we also tried to check their work by visiting the sites to see if they were outdated. We ended up with a list of 30 websites. Then, using the Python package Twarc, we put a search request to the Twitter API to find tweets that shared links from those websites. By searching on the websites’ home urls (for example, americannews.com) we were able to find any tweets that shared either the websites’ home pages or articles from those websites.

Second, we processed and cleaned the data to make it readable and usable. This required several steps. First we loaded the results of our search into a Pandas dataframe. Then we filtered out tweets that were retweets; while we think users who retweet fake news would be valid targets for our study, when we attempted to reply to a retweeted status, the API could not find the user who was retweeting the status. We decided that having our reply directly below the tweet sharing fake news was essential for our project, so we removed retweets from our collection. Next we processed the screen name and website name that the user was sharing: these two fields were used in our survey when responding to people. We also found there were some users who tweeted multiple links from the same website; we reasoned that surveying them multiple times about the same site would not be helpful, so we removed those duplicates. Finally we created a new data frame with a smaller sample of tweets that would be our first round of tweets to reply to. The new dataframe contained 720 tweets randomly drawn from our overall sample because we planned to send out 1 tweet every 2 minutes for a full 24 hours.

Sending the tweets was our final step. The wording of our message was “Hi {Username}, how reliable do you feel {Website name} is on a scale of 1-5?”, and we filled in the username and website with information from our dataframe. After tweeting, our program would wait a random amount of time between 90 and 150 seconds, for an average of 2 minutes of rest. The account we tweeted from was @NewsSurveyBot.

### **Results:**

We were able to send 360 tweets from our account before getting blacklisted by Twitter, at which point we were not able to write tweets via the API. Our replies only gathered 3 responses, printed in the table below:

User handle (User name)	Tweet text
@LastWave2014 (AngryAmericansUnited)	more trustworthy than CNN, MSNBC and ABC.. and that's all that really counts

@DuitsP (Ina Makoce)	Sometimes they are very reliable, 3
@Pinch0salt	1

**Table 1: Our first set of responses**

Two out of three responses did respond via a number; we planned, had we gotten more responses, to use regular expressions to automatically parse the tweet text for numbers and compile those into a dataframe.

After being shut down by Twitter on May 3rd, we decided to try to create another account and use the same script. This account, @Leah12Zhang varied in some important ways. The persona for our first account tried to be very official: the icon was a logo from the Merrill College of Journalism, and the word “Bot” was in both the username and handle in order to clearly broadcast our intentions. We believe this was not effective in gathering responses, however, and was partially responsible for our < 1% response rate.



**UMDJournoBot**  
@NewsSurveyBot

Hi, I'm a bot run by researchers from the Philip Merrill College of Journalism at UMD. I'm surveying people about sharing "fake news".

📍 College Park, MD  
🔗 [cutcaster.com/photo/10013754...](https://cutcaster.com/photo/10013754...)  
📅 Joined April 2017  
👶 Born on January 1, 2004



**Leah**  
@Leah12Zhang

I'm a student journalist in University of Maryland working actively on social media and want to detect the trend of fake news spreading. Please help on my study

📍 Maryland, USA  
📅 Joined May 2017  
👶 Born on January 17, 1994

**Figure 1: Our two Twitter personas**

On our second round of designing a Twitter persona, we decided to obscure the fact that it was a bot. We used a real picture of Xinyun and said we were a “student journalist” (not a team of

students). The bot also sent out a more inflammatory message: “Hi {Username}, {Website} is fake news. Agree or disagree on a scale of 1-5 (1 disagree, 5 agree)”. This bot only tweeted 34 times before getting shutdown by Twitter as well. However, the different persona seemingly was effective because it also received three responses.

User handle (User name)	Tweet text
@kristinemangan1 (Kristine Mangan)	5
@90NutrientsADay (Michael MD)	I say a 4 (=>80% Fake)
@KosloffM (Vance M Kosloff)	They want to rewrite traditional American values and destroy the nuclear family

**Figure 2: Our second round of responses**

Two of these responses contained a number, and both, perhaps surprisingly, agreed that the news they were sharing was fake. The other responded in a series of seven tweets, the first of which is printed above. The user said “the lame stream media has an agenda but it’s not all a lie”, making it unclear if he was referring to the fake news site he shared. These responses also raise the question of what kind of response bias we might encounter in our survey; we might hypothesize that people who are the most invested in fake news are especially distrustful of the traditional media, so would be less likely to respond to a survey from a journalist. This would bias the results towards those who know the news is fake.

### **Analysis / Recommendations**

In the end, we were not able to answer our question about the motivations of those who share fake news. The relative diversity of responses to our question, however, suggests that there are many different reasons users on Twitter share fake news, so we believe that further research in this area should be conducted. The user @kristinemangan1 even deleted her first tweet after we responded to her, suggesting that there is some percentage of the population who shares fake news that will actually respond positively to being called out for it.

As we’ve stated, a major obstacle to our project was getting shut down by Twitter. In their guidelines for use, Twitter says to not “Spam or bother users, or otherwise send them unsolicited messages”. However, we found that there is some wiggle room within this guideline. Seeing as sending people unsolicited messages is often times an integral part of journalism, we want to offer some recommendations to avoid being shut down by Twitter.

#### **Guidelines for using Twitter programmatically:**

Things you can do to (try) to not get shut down by Twitter:

- Don’t send the same exact thing

- In our case we varied the person we wrote to and the name of the news source
- If you can write multiple versions of the same tweet, this may be helpful. If you have a survey, however, this is a little more difficult because the language must be precise.
- Take your time sending out tweets, and vary the time.
  - So, for example, maybe put 2 minutes - 2:30 minutes

Things you can do to (try) to maximize responses:

- Have a compelling message and target a niche audience who might want to respond
- Use an account that has some followers and isn't especially made for surveying people

Overall, there's a small bandwidth of tasks that are not so large that your account will get shut down but that require automation. One example is tasks that can occur over a long time; we had to stick with a 2 minute per tweet timeline in order to aim for a sizable sample of responses, but if we could have let our bot run over a month we could have kept a slower pace.

### **Conclusion:**

This project was an exploration into what parts of journalism can be automated: what is ethical and legal (ie. within Twitter's guidelines) and what works on a social level (getting responses to our survey). In response to a very widespread problem like fake news, we aimed to automate responses to people in order to reach a large audience. Unfortunately, Twitter considers this spam, and our target population was likely hostile to allegations of fake news. We learned there are some techniques you can use to maximize the tweets you can send before Twitter suspends you, but ultimately they have complete editorial control over their platform.

The project was valuable in that it provided an opportunity to develop, improve, and implement an idea via programming. Our bot only utilized the Pandas, Tweepy, and Twarc packages but in using those we deepened our knowledge of interacting with the Twitter API. Design choices for the bot were made iteratively; for example we manually searched for a few tweets and send them out to users, and in doing this experiment we were able to refine the language and twitter persona we were going to use. The difficulties we found in interacting with Twitter made our project frustrating at times but ultimately taught us a lot about challenges of doing computational journalism. We learned that automating replies on Twitter is almost an art in that there are subtle techniques in order to evade getting shut down by Twitter, and this represents an opportunity to improve on our current process. Working as a team was enjoyable and our strengths complemented each other well. Elliot wrote a majority of the code for the project, having worked with the Twitter API before. Rebecca wrote the introductory materials in this report and our original proposal. Xinyun, in addition to collaborating on the coding, completely oversaw the second run of our bot. Each team member designed the bot, created slides for our presentations, and presented our work equally.