Maybe Scrolly?

7,888 views

Kristen Ruby

Follow @sparklingruby 33.1K followers

Dec 12 • 34 tweets • 11 min read

Bookmark Save as PDF + My Authors

BREAKING: US Political Misinfo Twitter Detection AI List

Thread Below ⬇️

1. I have obtained exclusive access to documents that have never been publicly revealed. The information I am about to share with you has never been released. This information is of public concern and the below thread includes my reporting on it.

2. I have thought long and hard about whether I should release this information. Ultimately, the public has a right to this information. Therefore, I am releasing it to you, despite my deep concerns about what this means and the possible implications.

3. Unlike the other writers involved in reporting on The Twitter Files, I do not have access to any direct information from Twitter. My information is from a direct source. I do not have the same level of legal protection these other writers have. I am taking the risk anyway.

4. I have never been focused on people or personalities. That type of reporting bores me and does nothing to help people understand the truth. I am focused on data.Therefore, decided to find the data on my own. What I found was shocking.

5. I have received two separate files with thousands of data cells. I will only be releasing a small portion of it and still have a lot to go through. It is important to understand the context of the information.

6. I will not be sharing the tweets that accompany the detection / flagging list for the parameters I am about to share to protect the privacy of those mentioned in them and the original accounts who posted the information.

7. Not all of the tweets flagged were removed. Some are still up. But what is important to understand is that they were flagged at all. Why did certain words trigger the algorithm? Why were these terms being actively monitored?

9. Do not share my reporting or screenshots without properly citing me as the author. All of this will be on Blockchain to show proof of authorship and ownership as well- so don't even try.

10. What was Twitter searching for?

"These are tweets we flagged for removal algorithmically. Some were held up, some reversed. We also searched for context."

11. "These are phrases we judged when in the right context were indicators of misinfo. Context was important because a journalist could tweet these & it would be okay. One or more of them was fine. In combination & in context, they were used to judge the tweet as misinformation."

12. To the left of the detection box you can see the flagged tweeted along with an algorithm and a score. To the right, you can see the rating, annotation, and decision. Not every auto detection of these words resulted in removal. Some stayed up. Some were removed.

12. The real question is: why were the algorithms flagging these phrases and words? These terms were flagged under US political misinformation. Below please find examples from the dataset I received. The world has never seen this data. I am not sure if Musk has seen this either.

13. I will not be sharing the example of the tweets that flagged these detections. I am only sharing the words parsed in the detection list.

Final reminder: not every tweet was removed.

Why were these terms monitored by Twitter as US political misinformation?

14. These are the *real* Twitter Files. Not curated internal Slack messages. And not screenshots that Twitter executives have shared with me. #RubyFiles

15. #RubyFiles

H

detections

```
[{'name': 'ballot', 'detections': array(['vote', 'ballot'], dtype=object)}
 {'name': 'election_dates', 'detections': array(['election day'], dtype=object)}
 {'name': 'gov_agency', 'detections': array(['ice'], dtype=object)}
 {'name': 'insult', 'detections': array(['corrupt'], dtype=object)}
 {'name': 'issue', 'detections': array(['corrupt', 'cheat', 'fak', 'stuff', 'make up'], dtype=object)}
 {'name': 'machine', 'detections': array(['ems', 'print'], dtype=object)}
 {'name': 'must_have', 'detections': array(['corrupt', 'dems', 'vote', 'win unless they cheat', 'cheat',
     'voting', 'ballot', 'stuff in the ballot', 'election'],
    dtype=object)}
 {'name': 'org', 'detections': array(['ice'], dtype=object)}
 {'name': 'participant', 'detections': array(['dems', 'voting'], dtype=object)}
 {'name': 'politics', 'detections': array(['ice', 'corrupt', 'vot', 'cheat', 'allot', 'election'],
    dtype=object)}
 {'name': 'post', 'detections': array(['boxes'], dtype=object)}
 {'name': 'result', 'detections': array(['win'], dtype=object)}
 {'name': 'tops', 'detections': array(['cheat', 'stuff in the ballot'], dtype=object)}
 {'name': 'various', 'detections': array(['win unless they cheat', 'stuff in the ballot'], dtype=object)}]
```

16. #RubyFiles

H

detections

```
[{'name': 'ballot', 'detections': array(['vote', 'ballot'], dtype=object)}
 {'name': 'election_dates', 'detections': array(['election day'], dtype=object)}
 {'name': 'gov_agency', 'detections': array(['ice'], dtype=object)}
 {'name': 'insult', 'detections': array(['corrupt'], dtype=object)}
 {'name': 'issue', 'detections': array(['corrupt', 'cheat', 'fak', 'stuff', 'make up'], dtype=object)}
 {'name': 'machine', 'detections': array(['ems', 'print'], dtype=object)}
 {'name': 'must_have', 'detections': array(['corrupt', 'dems', 'vote', 'win unless they cheat', 'cheat',
     'voting', 'ballot', 'stuff in the ballot', 'election'],
    dtype=object)}
 {'name': 'org', 'detections': array(['ice'], dtype=object)}
 {'name': 'participant', 'detections': array(['dems', 'voting'], dtype=object)}
 {'name': 'politics', 'detections': array(['ice', 'corrupt', 'vot', 'cheat', 'allot', 'election'],
    dtype=object)}
 {'name': 'post', 'detections': array(['boxes'], dtype=object)}
 {'name': 'result', 'detections': array(['win'], dtype=object)}
 {'name': 'tops', 'detections': array(['cheat', 'stuff in the ballot'], dtype=object)}
 {'name': 'various', 'detections': array(['win unless they cheat', 'stuff in the ballot'], dtype=object)}]
```

17. #RubyFiles

detections

```
ballot: [vote,ballot]
election_dates: [election day]
gov_agency: [ice]
insult: [corrupt]
issue: [corrupt,cheat,fak,stuff,make up]
machine: [ems,print]
must_have: [corrupt,dems,vote,win unless they cheat,cheat,voting,ballot,stuff in the
ballot,election]
org: [ice]
participant: [dems,voting]
politics: [ice,corrupt,vot,cheat,allot,election]
post: [boxes]
result: [win]
tops: [cheat,stuff in the ballot]
various: [win unless they cheat,stuff in the ballot]
```

18. #RubyFiles

detections

```
[{'name': 'ballot', 'detections': array(['ballot'], dtype=object)}
 {'name': 'counting', 'detections': array(['rested by fbi on 12 counts'], dtype=object)}
 {'name': 'gov_agency', 'detections': array(['fbi'], dtype=object)}
 {'name': 'issue', 'detections': array(['fraud', 'scheme'], dtype=object)}
 {'name': 'must_have', 'detections': array(['gop', 'republic', 'election official',
     'rested by fbi on 12 counts', 'voter fraud', 'election', 'ballot'],
    dtype=object)}
 {'name': 'org', 'detections': array(['fbi', 'elections board'], dtype=object)}
 {'name': 'participant', 'detections': array(['gop', 'republican', 'election official', 'voter'],
dtype=object)}
 {'name': 'politics', 'detections': array(['republic', 'election', 'arrest', 'vot', 'fraud', 'board',
     'commission', 'allot', 'schem'], dtype=object)}
 {'name': 'q', 'detections': array(['maga'], dtype=object)}
 {'name': 'tops', 'detections': array(['fraud'], dtype=object)}
 {'name': 'various', 'detections': array(['counts of voter fraud'], dtype=object)}
 {'name': 'worker', 'detections': array(['official', 'commissioner'], dtype=object)}]
```

19. #RubyFiles

detections

```
[{'name': 'ballot', 'detections': array(['votes', 'ballot'], dtype=object)}
 {'name': 'issue', 'detections': array(['prevent', 'fraud', 'accept', 'loss', 'stuff'], dtype=object)}
 {'name': 'must_have', 'detections': array(['votes', 'voter fraud', 'election', 'hillary', 'clinton', 'dem',
     'congress', 'stuffing ballot'], dtype=object)}
 {'name': 'org', 'detections': array(['congress'], dtype=object)}
 {'name': 'participant', 'detections': array(['voter', 'dem', 'congress'], dtype=object)}
 {'name': 'place', 'detections': array(['ave'], dtype=object)}
 {'name': 'politician', 'detections': array(['hillary', 'clinton'], dtype=object)}
 {'name': 'politics', 'detections': array(['vot', 'fraud', 'election', 'congress', 'arrest', 'allot'],
    dtype=object)}
 {'name': 'post', 'detections': array(['boxes'], dtype=object)}
 {'name': 'q', 'detections': array(['hillary'], dtype=object)}
 {'name': 'result', 'detections': array(['loss'], dtype=object)}
 {'name': 'tops', 'detections': array(['fraud', 'stuffing ballot'], dtype=object)}
 {'name': 'various', 'detections': array(['voter fraud', 'stuffing ballot'], dtype=object)}]
```

20. #RubyFiles

detections

```
[{'name': 'ballot', 'detections': array(['ballot'], dtype=object)}
 {'name': 'challenge', 'detections': array(['illegal'], dtype=object)}
 {'name': 'gov_agency', 'detections': array(['ice'], dtype=object)}
 {'name': 'insult', 'detections': array(['rats'], dtype=object)}
 {'name': 'issue', 'detections': array(['trust', 'hidden', 'fraud', 'harvest', 'illegal'], dtype=object)}
 {'name': 'must_have', 'detections': array(['trump', 'voter fraud', 'election fraud', 'democrat', 'ballot',
     'illegal'], dtype=object)}
 {'name': 'org', 'detections': array(['ice', 'state'], dtype=object)}
 {'name': 'participant', 'detections': array(['voter', 'democrat'], dtype=object)}
 {'name': 'politician', 'detections': array(['trump'], dtype=object)}
 {'name': 'politics', 'detections': array(['trust', 'hidden', 'vot', 'fraud', 'election', 'democra', 'ice',
     'allot', 'illegal', 'state'], dtype=object)}
 {'name': 'tops', 'detections': array(['fraud'], dtype=object)}
 {'name': 'unvoter', 'detections': array(['illegal'], dtype=object)}
 {'name': 'various', 'detections': array(['voter fraud', 'election fraud'], dtype=object)}]
```

21. #RubyFiles

detections

```
[{'name': 'ballot', 'detections': array(['votes', 'ballot'], dtype=object)}
 {'name': 'counting', 'detections': array(['counting vote'], dtype=object)}
```

```
{'name': 'hga', 'detections': array(['14'], dtype=object)}
{'name': 'issue', 'detections': array(['cheat', 'scam', 'harvest', 'unsecure'], dtype=object)}
{'name': 'must_have', 'detections': array(['gop', 'cheating scam november 2020', 're votes', 'obama',
    'ballot', 'voting', 'counting vote', 'blue state'], dtype=object)}
{'name': 'org', 'detections': array(['blue state'], dtype=object)}
{'name': 'participant', 'detections': array(['gop', 'voting'], dtype=object)}
{'name': 'politician', 'detections': array(['obama'], dtype=object)}
{'name': 'politics', 'detections': array(['cheat', 'scam', 'vot', 'campaign', 'allot', 'board', 'blue state',
    'secur'], dtype=object)}
{'name': 'post', 'detections': array(['boxes'], dtype=object)}
{'name': 'tops', 'detections': array(['cheat'], dtype=object)}
{'name': 'various', 'detections': array(['cheating scam november 2020', 'ballots in unsecur'],
dtype=object)}
{'name': 'watch', 'detections': array(['watch'], dtype=object)}]
```

22. #RubyFiles

detections
ballot: [ballot]
hga: [patriot]
issue: [stuff,fraud,hack]
machine: [comput]
must_have: [patriot,stuffed ballot,fraud,election]
org: [party]
participant: [patriot,party]
politics: [patriot,party,allot,commit,fraud,election,report]
post: [boxes]
tops: [stuffed ballot,fraud]
various: [stuffed ballot]

23. #RubyFiles

detections
ballot: [ballot]
challenge: [illegal]
issue: [prevent,fraud,stuff,illegal]
machine: [coding,detect]
must_have: [gov,gop,polling,voter fraud,ballot,stuffing of illegal ballot]
org: [federal,gov]
participant: [gov,gop,voter]
place: [station,ave]
politics: [federal,govern,monitor,vot,fraud,allot,illegal]
tops: [fraud,stuffing of illegal ballot]
unvoter: [illegal]
various: [voter fraud,stuffing of illegal ballot]
watch: [monitor]
worker: [monitor]

24. #RubyFiles

detections
ballot: [ballot]
challenge: [investigat,illegal]
gov_agency: [ice]
issue: [investigat,fraud,stuff,illegal]
must_have: [investigat,election fraud,operative,illegal,ballot,biden]
org: [ice,office]
participant: [person,police,operative]
politician: [biden]
politics: [sheriff,investigat,election,fraud,polic,operative,illegal,allot,office]
post: [boxes]
tops: [fraud]
unvoter: [illegal]
various: [election fraud]
watch: [investigat]

25. #RubyFiles

detections

```
[{'name': 'ballot', 'detections': array(['ballot'], dtype=object)}
 {'name': 'challenge', 'detections': array(['investigat', 'alleg', 'eviden'], dtype=object)}
 {'name': 'gov_agency', 'detections': array(['homeland security'], dtype=object)}
 {'name': 'issue', 'detections': array(['denie', 'investigat', 'fraud'], dtype=object)}
 {'name': 'must_have', 'detections': array(['election', 'investigat', 'ballot', 'homeland security',
'fraud'],
      dtype=object)}
 {'name': 'org', 'detections': array(['homeland security', 'doj'], dtype=object)}
 {'name': 'politics', 'detections': array(['election', 'investigat', 'allegation', 'allot', 'homeland',
       'secur', 'evidence', 'fraud'], dtype=object)}
 {'name': 'result', 'detections': array(['outcome'], dtype=object)}
 {'name': 'tops', 'detections': array(['fraud'], dtype=object)}
 {'name': 'various', 'detections': array(['fraud affecting 2020'], dtype=object)}
 {'name': 'watch', 'detections': array(['investigat'], dtype=object)}]
```

26. #RubyFiles

detections

challenge: [investigat]
counting: [data]
gov_agency: [fbi]
issue: [investigat,fraud,blackmail]
must_have: [investigat,district,election fraud,election]
org: [fbi,district]
participant: [attorney,person]
place: [ave]
politics: [ally,investigat,info,district,attorney,king,election,fraud,blackmail]
post: [mail]
tops: [fraud]
various: [election fraud]
watch: [investigat]
worker: [attorney,worker]

27. #RubyFiles

detections

```
[{'name': 'counting', 'detections': array(['counting'], dtype=object)}
 {'name': 'gov_agency', 'detections': array(['ice'], dtype=object)}
 {'name': 'issue', 'detections': array(['steal', 'believ', 'control', 'fool'], dtype=object)}
 {'name': 'machine', 'detections': array(['machine'], dtype=object)}
 {'name': 'must_have', 'detections': array(['stop the steal', 'trying to steal the elect'], dtype=object)}
 {'name': 'org', 'detections': array(['office'], dtype=object)}
 {'name': 'participant', 'detections': array(['activist'], dtype=object)}
 {'name': 'place', 'detections': array(['ave', 'center'], dtype=object)}
 {'name': 'politics', 'detections': array(['activist', 'office', 'ngo', 'election', 'control'], dtype=object)}
 {'name': 'tops', 'detections': array(['steal'], dtype=object)}
 {'name': 'various', 'detections': array(['stop the steal', 'trying to steal the elect'], dtype=object)}]
```

28. #RubyFiles

detections

challenge: [court]
issue: [change,cheat,fraud,changed]
machine: [ems]
must_have: [house and senate,dems,laws,cheat,election fraud,trump,voter]
org: [house,senate,laws,america]
participant: [supreme court,senate,dems,america,voter]
politician: [trump]
politics: [suprem,court,senat,laws,cheat,america,election,fraud,vot]
threat: [court,hanged]
tops: [cheat,fraud]
various: [election fraud,trump2024]

29. #RubyFiles

detections

```
[{'name': 'challenge', 'detections': array(['illegal'], dtype=object)}
 {'name': 'issue', 'detections': array(['vulnerab', 'access', 'illegal', 'lying', 'pressur', 'hack'],
      dtype=object)}
 {'name': 'machine', 'detections': array(['ems'], dtype=object)}
 {'name': 'must_have', 'detections': array(['gov', 'dominion', 'voting', 'my pillow', 'illegal',
       'election official', 'election', 'voting systems'], dtype=object)}
```

```
{'name': 'org', 'detections': array(['gov'], dtype=object)}
{'name': 'participant', 'detections': array(['gov', 'voting', 'election official'], dtype=object)}
{'name': 'politics', 'detections': array(['gov', 'vot', 'illegal', 'pressur', 'election'], dtype=object)}
{'name': 'q', 'detections': array(['maga'], dtype=object)}
{'name': 'unvoter', 'detections': array(['illegal'], dtype=object)}
{'name': 'various', 'detections': array(['my pillow'], dtype=object)}
{'name': 'worker', 'detections': array(['official'], dtype=object)}]
```

30. #RubyFiles

```
detections
[{'name': 'ballot', 'detections': array(['ballot'], dtype=object)}
 {'name': 'gov_agency', 'detections': array(['fda'], dtype=object)}
 {'name': 'issue', 'detections': array(['fraud', 'crime', 'traffick'], dtype=object)}
 {'name': 'must_have', 'detections': array(['election fraud', 'ballot'], dtype=object)}
 {'name': 'org', 'detections': array(['fda'], dtype=object)}
 {'name': 'politics', 'detections': array(['pm', 'election', 'fraud', 'crime', 'prosecut', 'allot', 'king'],
    dtype=object)}
 {'name': 'tops', 'detections': array(['fraud'], dtype=object)}
 {'name': 'various', 'detections': array(['election fraud'], dtype=object)}]
```

31. #RubyFiles

```
detections
ballot: [ballot]
issue: [trick,rig,change]
must_have: [antifa,operative,ballot,rig]
participant: [antifa,operative,people]
partisan_media: [gab]
place: [ave]
politician: [abbott]
politics: [antifa,operative,ally,allot,people,vot]
post: [boxes]
tops: [rig]
```

32. #RubyFiles

```
detections
[{'name': 'ballot', 'detections': array(['votes'], dtype=object)}
 {'name': 'issue', 'detections': array(['stolen', 'fraud'], dtype=object)}
 {'name': 'must_have', 'detections': array(['stolen', 'election', 'president', 'election fraud', 're votes'],
    dtype=object)}
 {'name': 'participant', 'detections': array(['president'], dtype=object)}
 {'name': 'partisan_media', 'detections': array(['newsmax'], dtype=object)}
 {'name': 'place', 'detections': array(['ave', 'place'], dtype=object)}
 {'name': 'politics', 'detections': array(['election', 'presid', 'fraud', 'vot'], dtype=object)}
 {'name': 'post', 'detections': array(['boxes', 'box'], dtype=object)}
 {'name': 'result', 'detections': array(['outcome'], dtype=object)}
 {'name': 'tops', 'detections': array(['stolen', 'fraud'], dtype=object)}
 {'name': 'unvoter', 'detections': array(['resident'], dtype=object)}
 {'name': 'various', 'detections': array(['election fraud'], dtype=object)}]
```

33. #RubyFiles

```
detections
challenge: [eviden]
insult: [rat]
issue: [fraud,infiltrat]
machine: [dre]
must_have: [gov,election official,election fraud,election,democrat]
org: [gov,state]
participant: [gov,election official,democrat]
partisan_media: [telegram]
politics: [gov,state,election,evidence,fraud,democra]
tops: [fraud]
various: [election fraud]
worker: [official]
```

• • •

Missing some Tweet in this thread? You can try to force a refresh

Tweet Share Email

Keep Current with Kristen Ruby

Stay in touch and get notified when new unrolls are available from this author!

+ Add to "My Authors" Read all threads

This Thread may be Removed Anytime!

Twitter may remove this content at anytime! Save it as PDF for later use!

Save this thread as PDF

People who liked this thread also liked...

Bari Weiss
@bariweiss

Dec 9

THREAD: THE TWITTER FILES PART TWO. TWITTER'S SECRET BLACKLISTS.

1. A new #TwitterFiles investigation reveals that teams of Twitter employees build blacklists, prevent disfavored tweets from trending, and actively limit the visibility of entire accounts or even trending topics—all in secret, without

Read 31 tweets

Matt Taibbi
@mtaibbi

Dec 9

1. THREAD: The Twitter Files THE REMOVAL OF DONALD TRUMP Part One: October 2020-January 6th

2. The world knows much of the story of what happened between riots at the Capitol on January 6th, and the removal of President Donald Trump from Twitter on January 8th...

Read 65 tweets

Michael Shellenberger
@ShellenbergerMD

Dec 10

1. TWITTER FILES, PART 4 The Removal of Donald Trump: January 7 As the pressure builds, Twitter executives build the case for a permanent ban

On Jan 7, senior Twitter execs: - create justifications to ban Trump - seek a change of policy for Trump alone, distinct from other political leaders -

Read 40 tweets

Try unrolling a thread yourself!

1. Follow @ThreadReaderApp to mention us!

2. From a Twitter thread mention us with a keyword "unroll"

```
@threadreaderapp unroll
```

Practice here first or read more on our help page!

More from @sparklingruby

Kristen Ruby
@sparklingruby

Dec 15

BREAKING: An inside look at the type of content Twitter flagged for review Thread below ⬇️

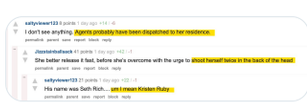
1. "This file is international. The other file was US only." The following screenshots will reveal the type of content Twitter was monitoring for political misinformation. The file included dates, language & annotation.

Read 24 tweets

Kristen Ruby
@sparklingruby

Dec 13

What happens when people report the truth?

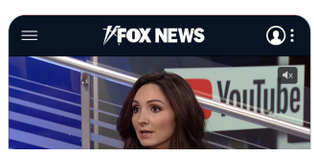


Read 4 tweets

Kristen Ruby
@sparklingruby

Dec 13

2019: ⚡Twitter incorrectly flags tweets for 'sensitive content' "Do you want to hire more moderators or do you want to keep teaching the machines and having them learn on the job?"



Read 4 tweets

Kristen Ruby
@sparklingruby

Dec 9

BREAKING: Former Twitter employee shares exclusive details with me on AI, Access to DMs, and more. Thread below ⬇️

1. What is guano?

[redacted] So every database serves a purpose. Tracking actions or tweets etc etc. we had probably 100s if not

Read 36 tweets

Kristen Ruby
@sparklingruby

Dec 7

#ChatGPT can write very good obituaries. Pro tip: use the output to update your executive bio. The AI obit is written in third person. Example 👇🏻

"Kristen Ruby is from Scarsdale, New York. Kristen is the founder of @RubyMediaGroup, an award-winning social media agency. Ruby is known for her creative and highly effective

Read 4 tweets

Did Thread Reader help you today?

Support us! We are indie developers!

This site is made by just two indie developers on a laptop doing marketing, support and development! Read more about the story.

Become a Premium Member ($3/month or $30/year) and get exclusive features!

Become Premium

Don't want to be a Premium member but still want to support us?

Make a small donation by buying us coffee ($5) or help with server cost ($10)

Donate via Paypal

Or Donate anonymously using crypto!

Ethereum

0xfe58350B80634f60Fa6Dc149a72b4DFbc17D341E copy

Bitcoin

3ATGMxNzCUFzxpMCHL5sWSt4DVtS8UqXpi copy

Thank you for your support!

Help | About | TOS | Privacy

Follow Us on Twitter!

Tweet Share