SCHOOL OF INFORMATION CENTER FOR
SOCIAL MEDIA RESPONSIBILITY
UNIVERSITY OF MICHIGAN

Iffy Quotient: A Platform Health Metric for Misinformation

Paul Resnick[1], Aviv Ovadya[2], and Garlin Gilchrist[3]

v1: October 10, 2018 (link)
v2: July 23, 2019 (link)
v3 (this version): June 29, 2023 (link)
Latest version can always be accessed at http://umsi.info/iffy-quotient-whitepaper
Website with public dashboard: https://csmr.umich.edu/projects/iffy-quotient/

Executive Summary

Social media sites and search engines have become the de facto gatekeepers of public communication, a role once occupied by publishers and broadcasters. With this new role come public responsibilities, including limiting the spread of misinformation.

Externally maintained metrics offer a way to measure the progress of media platforms at meeting their public responsibilities. By contrast with the current environment of accountability by anecdotes, Platform Health Metrics can focus attention on the overall performance of platforms rather than on bad outcomes in individual cases.

The Center for Social Media Responsibility at the University of Michigan School of Information has developed the Iffy Quotient, a metric for how much content from “Iffy” sites has been amplified on Facebook and Twitter. We use the term “Iffy” to describe sites that frequently publish misinformation. It is a light-hearted way to acknowledge that our categorization of the sites is based on imprecise criteria and fallible human judgments. We are publishing a web-based dashboard that charts the Iffy Quotient since early 2016. The dashboard enables comparisons over time and between platforms. This report describes the calculation of the Iffy Quotient in detail, discusses some of its potential limitations, and analyzes some of the trends.

[1] Paul Resnick is a Professor and Associate Dean for Research and Innovation at the University of Michigan School of Information, and Director of the Center for Social Media Responsibility. He was a consultant to Facebook in 2018-2020, but this work is independent.
[2] Aviv Ovadya served as Chief Technologist for the Center for Social Media Responsibility in 2017-18, with primary responsibility for architecting the Iffy Quotient.
[3] Garlin Gilchrist helped to draft the initial version of this report while he was Executive Director of the Center for Social Media Responsibility in 2017-18. He currently serves as Lieutenant Governor of the State of Michigan.

Major Changes for v3

In 2023, for v3, we switched from calculating the fraction of URLs that were from Iffy sites to calculating the fraction of total engagement on those URLs that were from Iffy sites. We had previously displayed the engagement-weighted Iffy Quotient as an alternative metric. In 2023, we switched to make it our primary metric.

Second, we switched from calculating daily Iffy Quotients to weekly, with weeks running from Monday-Sunday. In our historical chart we were previously performing seven-day smoothing: each day's Iffy Quotient was really the Iffy Quotient over the previous seven days. The new interface for our web page, allowing for comparisons between time periods, came out cleaner when we just reported on weeks instead of seven-day moving averages. If the idea of seven-day moving averages seems complicated to you, then you are already appreciating why switching to reporting on the Iffy Quotient for each week was a desirable simplification!

Third, we changed our "backfill" practice. Previously, we used any new classifications or classification changes to retroactively update the Iffy Quotient calculations for the previous three months. The motivation for that practice was that newly popular sites might not be classified immediately by NewsGuard or Media Bias/Fact Check. The backfill allowed us to apply the classifications retroactively, once they came in. However, the choice of three months as the time period for backfill was somewhat arbitrary. Moreover, it meant that all Iffy Quotient values that we published were provisional for a period of three months. We no longer apply site classifications retroactively. Once we publish a chart with the Iffy Quotient for a time period, we will no longer change it (unless we discover a data error).

Finally, we applied a correction for near-duplicate URLs. Sometimes, among the top 5000 URLs returned by NewsWhip on a given day, multiple URLs differed only in the "query parameters," the part of the URL path after the "?" character. For some sites, URL pairs differing only in these query parameters reflected different news stories. In other cases, however, they reflected a single news story that had been retrieved by NewsWhip multiple times, with growing engagement counts over time but slightly different URLs. Through most of our data collection, this happened rarely, but there was a time period when it happened more frequently. We applied a correction to account for these near-duplicate URLs. Details are described below in Step 1 (data collection) of our calculation methodology. Compared to not correcting at all, the main effect of noticing and correcting for near-duplicate URLs was to reduce the engagement attributed to sites unlabeled by either NewsGuard or MBFC (e.g., nba.com), and thus it had only small effects on the calculated Iffy Quotient.

Major Changes for v2

In 2019, we changed the methodology for calculating the Iffy Quotient. We now deem a site as Iffy or OK based on ratings from NewsGuard, falling back on ratings from Media Bias/Fact Check only for sites not rated by NewsGuard. In addition, because NewsGuard has higher coverage of sites that were popular in 2019 than those in 2016, and for sites that were popular on Facebook than those that were popular on Twitter, we now define the Iffy Quotient as the fraction of URLs from rated sites rather than the fraction of URLs from all sites. This has increased the absolute values of the Iffy Quotient on particular days in comparison to our version 1 numbers, but the trends over time remain largely unchanged.

Report

Introduction

Social media sites and search engines have become the de facto gatekeepers of public communication, a role once occupied by publishers and broadcasters. With this new role come public responsibilities, beyond the commercial responsibilities that a company has to please customers and reward shareholders. Among these public responsibilities are limiting the spread of misinformation, ensuring a level playing field in the free competition of ideas, and promoting interpersonal connections that heal rather than aggravate societal divisions.

The Center for Social Media Responsibility is developing Platform Health Metrics, which track how well social media sites and search engines (which we refer to collectively as media platforms) are meeting these public responsibilities. A metric reduces an abstract ideal, such as limiting the spread of misinformation, to a concrete measurement that can be taken repeatedly, enabling comparisons over time and between platforms. This report introduces the Iffy Quotient, one such metric. It computes the fraction of the most popular URLs that come from Iffy sites - sites that have frequently published misinformation and hoaxes in the past. Our website reports the Iffy Quotient for Facebook and Twitter going back to 2016 and it will be updated on an ongoing basis.

> For both Twitter and Facebook, there was an increase in attention to Iffy sites in the runup to the 2016 U.S. elections. The Iffy Quotient nearly doubled on each site from April to November, peaking right around election day.

Key Performance Metrics are powerful management tools for media companies. Mature consumer-facing technology platforms already maintain internal suites of metrics, such as monthly page views, clickthrough rates, dwell times, customer acquisition and retention, and ad revenue. These metrics strongly influence decisions about changes to products and policies. Typically, product managers are rewarded for improving some primary metric, subject to the constraint that there is at most a modest decline in other metrics.

Externally maintained metrics offer two advantages over internal metrics maintained by the platforms. First, they can draw attention to issues that platforms may either not be tracking themselves or not prioritizing as much as the public would like. This form of public accountability is preferable to the current environment of accountability by “gotcha” anecdotes. It focuses attention on the overall performance of platforms rather than bad outcomes in individual cases; some bad outcomes may be inevitable given the scale on which the platforms operate.

Second, external metrics can create public legitimacy for claims that platforms make about how well they are meeting public responsibilities. Even if Facebook actually reduces the audience share for Iffy content, the public may be skeptical if Facebook defines the metric, conducts the measurement without audit, and chooses whether to report it.

Of course, metrics are not a panacea. The thing that is measured is often a proxy for the thing that really matters. Managerial efforts to improve the metric (the proxy) may not similarly improve the true quantity of interest; in extreme cases this is referred to as “gaming the metric.” Externally maintained metrics may be especially susceptible to such problems. Due to limited access to proprietary data, an external metric may involve compromises that make it a weaker proxy. This report also describes some of the limitations and compromises involved in defining and measuring the Iffy Quotient and the potential risks that come from them.

Limiting the Reach of Misinformation is a Public Responsibility of Media Platforms

Media platforms should not be expected to prevent the publication of misinformation or to prevent people who seek access to it from finding it (unless the misinformation is also harmful in some way that is prohibited by law, such as by directly inciting violence). We think, however, that media platforms should not amplify misinformation—as Renee Diresta concisely argues: “Free speech is not the same as free reach.”[4] Misinformation, if widely shared, can influence public opinion, create social divisions, and even stir up violence, as has been documented in India, Sri Lanka, and Myanmar.[5] It can degrade trust, which is necessary for society to function well. And it can drive government actions that benefit special interests rather than public interests. It is especially important that special interests, including foreign actors, not be able to manipulate media platforms so that they spread misinformation.

Facebook and Twitter have accepted responsibility for countering deliberate manipulation. For example, Twitter executive Colin Crowell wrote on the company blog in 2017, “We’re working hard to detect spammy behaviors at source, such as the mass distribution of Tweets or attempts to manipulate trending topics.”[6] Facebook reported that it removed 583 million fake accounts in the first quarter of 2018.[7] The two seem to diverge somewhat, however, in whether they accept responsibility for not amplifying misinformation in the absence of deliberate manipulation. Facebook appears to implicitly accept this responsibility, with their announcement that they will take action to reduce the audience for items that journalist fact-checkers judge to be false.[8] Crowell’s post, on the other hand, goes on to suggest that Twitter thinks its only responsibility is to ensure the spread of counter-information to misinformation.

[4] https://www.wired.com/story/free-speech-is-not-the-same-as-free-reach/

[5] https://www.nytimes.com/2018/07/18/technology/facebook-to-remove-misinformation-that-leads-to-violence.html

[6] https://blog.twitter.com/official/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html

[7] https://newsroom.fb.com/news/2018/05/enforcement-numbers/

[8] https://www.facebook.com/help/1952307158131536

Whether or not they accept responsibility for preventing the amplification of misinformation, executives from both sites argue that they should not become “arbiters of truth.” This still leaves room, however, for other kinds of counter-measures that could reduce the platforms’ amplification of misinformation. One approach focuses entirely on process, eschewing assessments of truth—for example, identifying and eliminating bot accounts. Another is to outsource judgments of information reliability to journalists, third-party fact-checkers, media watchdogs, or platform users. This report does not argue for or against any particular counter-measures; the Iffy Quotient merely provides a way to measure the effectiveness of any counter-measures that may have been implemented.

Defining Misinformation

Exactly what counts as misinformation has become politically contested and is an active front in the ongoing culture wars. Many seemingly factual claims are in fact contextual spin that may invite misinterpretation without actually making any factual claims. Even with factual claims, only limited evidence may be available. Not everyone may agree, given the available evidence, whether the claim is true or false.

We begin with a recipient-centric, subjective definition of misinformation: information that *the recipient would judge* to be false or misleading *if they took the time to carefully consider all of the evidence about it*. With that in mind, the platform’s responsibility, we argue, should be to spread a piece of information only to those people who would, in full knowledge of available evidence, access and spread it.

However, media platforms take many actions at an aggregate, non-personalized level. That makes it useful to adopt a collective, but still subjective, definition of misinformation: information that *most recipients would judge* to be false or misleading *after taking the time to carefully consider all of the evidence about it*. With that definition, the platform’s responsibility is to support the amplification of a piece of information if a majority of the potential audience, in full knowledge of available evidence, would access and spread it.

These definitions are based on a counterfactual: how individuals would judge an item *if* they considered the evidence for and against it. In an ideal world, there would be a process that resolved that counterfactual for enough people to yield a good estimate of the collective judgment. In the future, automated classifiers might be used as a proxy, with verification against human judgments on a sample of items. In current practice, we rely on external entities as a proxy to speak for what the collective judgment would be.

Site-level Proxy Judgments

In our case, we rely on two external entities, NewsGuard[9] and Media Bias/Fact Check.[10] They make judgments not on individual items but on entire sites. We treat their judgments as a proxy for whether a particular content site frequently publishes information that the majority of people would consider false or misleading, after considering all evidence. Because these external entities make their judgments based on imprecise criteria, we adopt the whimsical term “Iffy” rather than the more definitive term “misinformation.”

Site-level judgments alone would not be appropriate for platforms to use in making decisions about whether to amplify the audience for particular items, for two reasons. First, even sites that frequently publish misinformation may also publish some things that are not misleading. Second, the judgments made by NewsGuard and Media Bias/Fact Check, according to the criteria they define and articulate, may not match what the majority of people would consider false or misleading. This reasoning is evident in the announced practices of media platforms. For example, Facebook takes action to reduce the audience for items that journalist fact-checkers judge to be false,[11] but, to our knowledge, does not rely on the site-level judgments of NewsGuard or Media Bias/Fact Check.

It is reasonable, however, to use site-level judgments in calculating the Iffy Quotient. Treating all items from a site as Iffy will give an overestimate of the absolute amount of misinformation distributed by the platform. The amount of overestimation should be fairly stable, however, and thus should not affect comparisons between sites and comparisons over time. Similarly, mistaken judgments about whole sites could lead to errors in the absolute percentage of Iffy content as recorded in an Iffy Quotient measurement, but are unlikely to have a large effect on comparisons between measurements. Thus, the Iffy Quotient is suspect as a measure of the absolute amount of misinformation that is spread by platforms, but is a reasonable way to judge whether Twitter or Facebook spread more misinformation at a particular point in time, whether Twitter spread more or less Iffy content in July 2018 vs. July 2017, or whether a change in media platform moderation policy enacted on a particular date led to less amplification of Iffy content the following month.

Calculating the Iffy Quotient for English Language News

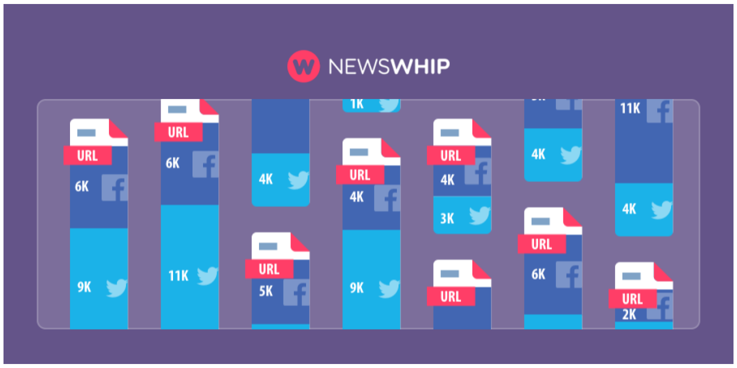
For each day, we download the most popular URLs on Facebook and Twitter from NewsWhip,[12] a commercial social media monitoring company. We also download site judgments from NewsGuard and Media Bias/Fact Check about which sites are Iffy. Finally, we calculate the fraction of engagement received by URLs from Iffy sites. Details follow.

[9] https://www.newsguardtech.com/
[10] https://mediabiasfactcheck.com
[11] https://www.facebook.com/help/1952307158131536
[12] https://www.newswhip.com

NewsWhip tracks the creation of URLs on more than 400,000 sites every day.

NewsWhip maintains a list of sites that it monitors. This includes traditional news sites, sites that people treat like news, and other sites that are relevant to NewsWhip's clients. This does not include all websites, but it is quite broad.

Limitation: It is possible that the incompleteness of NewsWhip's tracking could lead to incorrect estimation of the true Iffy Quotient. For example, if NewsWhip is less effective at discovering high-engagement fly-by-night sites (e.g., Macedonian traffic arbitrage) than it is at tracking established reliable sources, a higher fraction of the missed sites may be Iffy, leading to our measured value being an underestimate. Or it could go the other way, if NewsWhip misses more URLs from untracked reliable sites (perhaps small niche sites) than from Iffy sites.

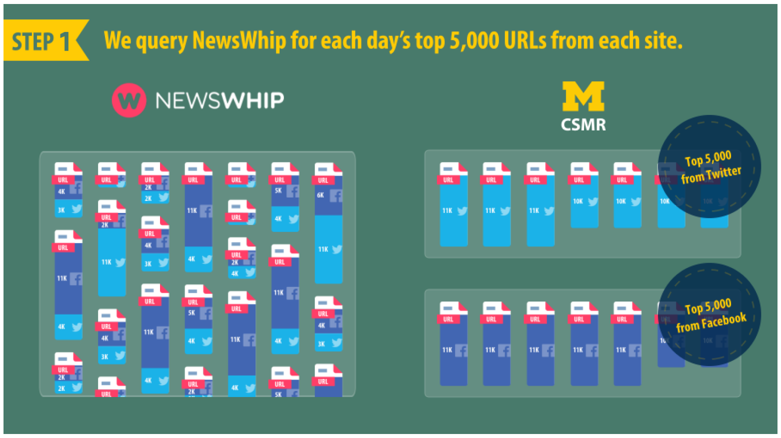


For each news URL, NewsWhip gathers engagement data on Facebook and Twitter.

NewsWhip tracks new URLs added to the sites it monitors. Whenever a new page is added to one of these sites, NewsWhip checks for social engagement with that URL on Facebook and Twitter as described below. NewsWhip also continues to check social engagement repeatedly, though more and more rarely as the number of engagements appears to stabilize.

NewsWhip provides a Facebook engagement score for each URL, which appears to closely track data available through the public Facebook Graph API. It provides an indicator of aggregate engagements (likes and other reactions, shares, comments) with a URL without revealing the identity of any particular person who engages with it. NewsWhip associates all engagement data with the URL and its publication date, regardless of when the engagement occurred on Facebook.

NewsWhip also provides a "Twitter Influencer Shares" value for URLs. This is the sum of tweets mentioning a URL by the "influencers" that NewsWhip tracks and the retweet counts for those tweets. NewsWhip tracks at least 300,000 accounts, including verified Twitter accounts and other accounts that are useful to NewsWhip’s clients. This approach has been used starting around November 27, 2017. Prior to that, NewsWhip used a different method to estimate the number of tweets and retweets. Exact numbers before and after the changeover may not be comparable. The change should, however, affect URLs from Iffy sites and other sites in a similar way, and thus should not appreciably affect the Iffy Quotient.


STEP 1
We query NewsWhip for each day’s top 5,000 URLs from each site.
NEWSWHIP
CSMR
Top 5,000 from Twitter
Top 5,000 from Facebook

Step 1. We query NewsWhip for each day’s top 5,000 URLs from each site.

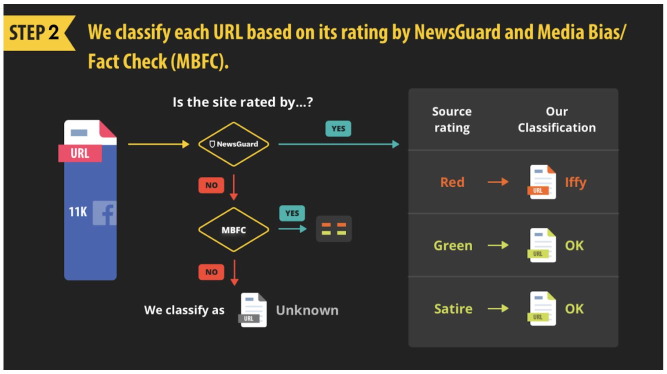
We query NewsWhip for 5,000 URLs published on each date. Since engagements are still trickling in for recently published content, we treat the engagement data queried two days later as the official engagement scores and do not update beyond two days.

We applied a correction to discard some near-duplicate URLs that differed only in their query parameters. In some cases, NewsWhip retrieved engaged counts multiple times and provided all of those engagement counts with slightly different URLs. Unfortunately, the information available to us daily from NewsWhip does not make it possible to distinguish these erroneous near-duplicates from genuine near-duplicates that refer to distinct news articles.

NewsWhip provided historical data that allowed us to distinguish retrospectively. We used that to make a determination for each domain of whether to treat near-duplicates as distinct, or whether to discard all but the last, highest engagement score. We first grouped potential near-duplicates based on their URL address, timestamp, headline, and summary. We then maintained a lookup table of domains to determine whether, in the future, to apply the correction of discarding all but the last near-duplicate URL in an identified group. In the historical data, some domains had fewer groups of near-duplicates URLs that reflected different underlying articles than groups that did not. We decided to apply the correction to those domains. We conducted sensitivity analysis and determined that, on our historical data, using the lookup table to determine whether to apply the correction on a domain-by-domain basis yielded very similar final Iffy Quotient estimates to what would have resulted from determining whether to apply the correction on a per-news-article basis.

Compared to not correcting at all, the main effect of noticing and correcting for near-duplicate URLs was to reduce the engagement attributed to sites unlabeled by either NewsGuard or MBFC (e.g., nba.com), and thus it had only small effects on the calculated Iffy Quotient.

Limitation: Fake accounts may be used to inflate the engagement counts on Twitter and Facebook, as a way to make them appear more popular and thus drive real traffic. Twitter and Facebook try to root out such fake accounts, and NewsWhip tries to avoid inclusion of fakes among the Twitter accounts that it tracks, but those attempts can never be completely successful. Fake accounts are perhaps more likely to be used for content from Iffy sites than other sites. This could lead to an overestimate of the true Iffy Quotient, by causing more Iffy sites to creep into the top 5,000 than would be there if no fake accounts were included in engagement scores.



STEP 2 We classify each URL based on its rating by NewsGuard and Media Bias/Fact Check (MBFC).

Is the site rated by...?

URL

11K

NewsGuard

YES

NO

MBFC

YES

NO

We classify as Unknown

Source rating	Our Classification
Red →	Iffy
Green →	OK
Satire →	OK

Step 2. We classify each URL based on its rating by NewsGuard and Media Bias/Fact Check.

Given a URL, we classify it by comparing the hostname to the hostnames listed in the *source lists* curated and maintained by NewsGuard and Media Bias/Fact Check. Starting with version 2, our primary source is now NewsGuard, because it is run by career journalists and is more transparent about its criteria and judgments of individual sites.[13] For sites not rated by NewsGuard, we fall back on Media Bias/Fact Check. In version 1, we reported the Iffy Quotient based on Media Bias/Fact Check only. Sites rated by neither NewsGuard nor Media Bias/Fact Check are treated as Unknown. We also treat as Unknown any URL from a platform site that does not publish or curate its own news content, because there is no single site-level rating that would meaningfully apply to all content on such sites. For versions 1 and 2, the platform list was YouTube and Facebook. For version 3, the list is: Google, Facebook, Instagram, Medium, Pinterest, Telegram, TikTok, Twitter, Vimeo, Weibo, WhatsApp, and YouTube.[14]

NewsGuard ratings are available for any website via a browser plugin. By agreement with NewsGuard, they provide us with updates of their complete list of site ratings.

[13] https://www.newsguardtech.com/about/why-should-you-trust-us/
https://www.newsguardtech.com/ratings/criteria-for-and-explanation-of-ratings/

[14] The version 3 exclusion list is borrowed from https://www.propublica.org/article/google-ads-misinformation-methodology and slightly amended.

NewsGuard's Reliability Ratings are based on nine apolitical and basic journalistic criteria that assess the credibility and transparency of a news or information site, including "Does not repeatedly publish false content," "Regularly corrects or clarifies errors," and "Avoids deceptive headlines." NewsGuard awards points for each criterion and sums them up; a score less than 60 earns a "generally unreliable" rating; 60 and above earns a "generally reliable" rating. They also identify some sites as "Satire". We treat sites with scores of 60 and above, or "satire" labels, as OK, and other sites with scores below 60 as Iffy.

Media Bias/Fact Check (hereafter "MBFC") evaluates sites and puts them on one or more of the lists, based on criteria they describe for each of the lists. We classify as Iffy any site that is on either the "Questionable Sources" list or the "Conspiracy-Pseudoscience" list. MBFC's website describes the criteria for each as follows:

> A questionable source exhibits *one or more* of the following: extreme bias, overt propaganda, poor or no sourcing to credible information and/or is fake news. Fake News is the *deliberate attempt* to publish hoaxes and/or disinformation for the purpose of profit or influence. Sources listed in the Questionable Category *may* be very untrustworthy and should be fact checked on a per article basis.
>
> Sources in the Conspiracy-Pseudoscience category *may* publish unverifiable information that is *not always* supported by evidence. These sources *may* be untrustworthy for credible/verifiable information, therefore fact checking and further investigation is recommended on a per article basis when obtaining information from these sources.

MBFC also provides explicit listings of sites it has evaluated and judged *not* to be appropriate for one of those lists. Other lists they provide include "Left Bias," "Left-Center Bias," "Least Biased," "Right-Center Bias," "Right Bias," "Pro-Science," and "Satire." We classify as "OK" any site that is not Iffy and is on one of these other lists. If a site is not on any of MBFC's lists, we classify it as Unknown.

Complete listing of the MBFC judgments is available on its website. To give a flavor for the judgments:

- Vox and Upworthy are classified as OK (Left Bias)
  but Learn Progress and Occupy Democrats are Iffy (Questionable Sources).
- Fox News and the Drudge Report are classified as OK (Right Bias)
  but Breitbart and TruthFeed are Iffy (Questionable Sources);

For the first version of this report, we included a robustness test using Open Sources. However, Open Sources has evaluated many fewer sites, and the last update prior to this report was April 28, 2017.[15] Beginning with version 2, we omit Open Sources entirely.

[15] https://github.com/BigMcLargeHuge/opensources/commits/master

Publishers sometimes change their domain names. This is especially true for Iffy publishers who may change domain names to get a fresh start with search engines and social media sites that may have started to demote or demonetize them based on prior complaints and investigations. NewsGuard and Media Bias/Fact Check may not always notice these name changes right away. Moreover, when they do, they may put the new site on their lists but remove the old sites.

To account for incompleteness of some of our classifications that might result from sites changing their domain names, from 2016-2018 we tested for automatic redirects and computed some inferred labels. In particular, if a URL published some time ago now redirects to a site that NewsGuard or MBFC has listed, we give it the same classification as the new site. In our first release, we also went the other direction, applying the old label to a new site that the old label redirects to; we discontinued that practice after finding that some Iffy sites have shut down and redirect to mainstream sites. After January 1, 2019, we discontinued redirect processing entirely, because we found that there were very few newly popular URLs that were getting inferred labels as a result of the redirect processing.

Step 3. Sum the engagement scores for Iffy URLs, and separately for OK URLs, across all days in the selected time period.

For versions 1 and 2, we counted the number of URLs from Iffy and OK sites. That yielded a count-based Iffy Quotient. It treats all URLs in the top 5000 for each day as equal, even if the top URL gets a lot more engagement than the bottom one.

For version 3, we switched to computing an engagement-weighted version of the Iffy Quotient. We add up the total engagement score of all URLs from Iffy sites. Separately, we add up the total engagement score of all URLs from OK sites.

Step 4. Compute the Iffy Quotient as $\frac{Iffy}{Iffy + OK}$.

We define the Iffy Quotient as the fraction of total engagement for our pool of URLs that went to Iffy sites. We calculate this separately for Facebook and Twitter. We compute this for each week (Monday-Sunday), and also for months and calendar years.

In versions 1 and 2, the *Iffy* and *OK* quantities in the quotient (fraction) were counts of URLs rather than sums of engagement scores. For version 3, the quantities are sums of engagement scores for the URLs. Arguably, this engagement-weighted version is a closer proxy for the real quantity of concern, the fraction of human attention to unreliable information. However, it depends more heavily on the accuracy of the engagement estimates than the count-based Iffy Quotient, which depends on the engagement estimates only to select the top 5,000.

In version 1, the Iffy Quotient was computed as $\frac{Iffy}{Iffy + OK + Unknown}$, including Unknown URLs in the denominator. Essentially, that treated all Unknown sites as if they were OK. NewsGuard has greater coverage of sites that were popular in 2019 than those that were popular in 2016, and greater coverage of sites with popular URLs on Facebook than Twitter. To facilitate comparison over time and between sites, version 2 expresses the Iffy Quotient as the fraction of rated sites. In absolute terms, the Iffy Quotient was higher in the second version than the first version. We caution readers that they still should not place too much emphasis on the absolute value of the Iffy Quotient but instead should use it to make comparisons over time and between platforms.

Limitation: Some of the Unknown URLs may be from sites that NewsGuard would judge as Iffy but hasn't gotten around to judging yet. Our current approach assumes that the unrated sites would be rated as Iffy in the same proportion as the rated ones. That is, if 10% of the URLs from rated sites are from Iffy sites, then 10% of the URLs from unrated sites are also from Iffy sites. If, for example, the unrated sites are disproportionately iffy, as NewsGuard rates a larger fraction of them the Iffy Quotient could go up over time without there being any change in the fraction of content actually from Iffy sites. There is no way to completely eliminate this possibility that improved measurement over time could make it look like there is a change in the Iffy Quotient. Our deep dive page shows for any time period the most popular URLs that are classified as Iffy, OK, and Unknown. Anecdotally, many of the popular Unknown URLs are stories about Korean pop music.

In summary, here's how we calculate the Iffy Quotient.

1. NewsWhip provides the 5,000 most engaged-with URLs each day, on Facebook and Twitter.
2. We define as Unclassified any URLs originating from platform sites such as Google and Facebook.
3. NewsGuard provides lists of domain names they have judged.
   a. We define as Iffy those sites that have a score below 60 and are not satire or a platform.
   b. We define as OK those sites that have scores of 60 and above or are satire.
4. For sites unrated by NewsGuard, we check whether Media Bias/Fact Check has labeled it.
   a. We define as Iffy those sites listed as Questionable Sources or Conspiracy/Pseudoscience
   b. We define as OK those sites listed in other categories, including Left Bias and Right Bias
   c. We define as Unclassified any URLs from all remaining sites.
5. For each day, we sum up the engagement with URLs classified as Iffy and OK. The Iffy Quotient is the fraction of user engagement $\frac{Iffy}{Iffy + OK}$.

Our main page reports the Iffy Quotients for Facebook and Twitter for the previous week. Our deep dive page lets anyone choose any week, month, quarter, or year, and allows for comparisons between time periods.

Analysis

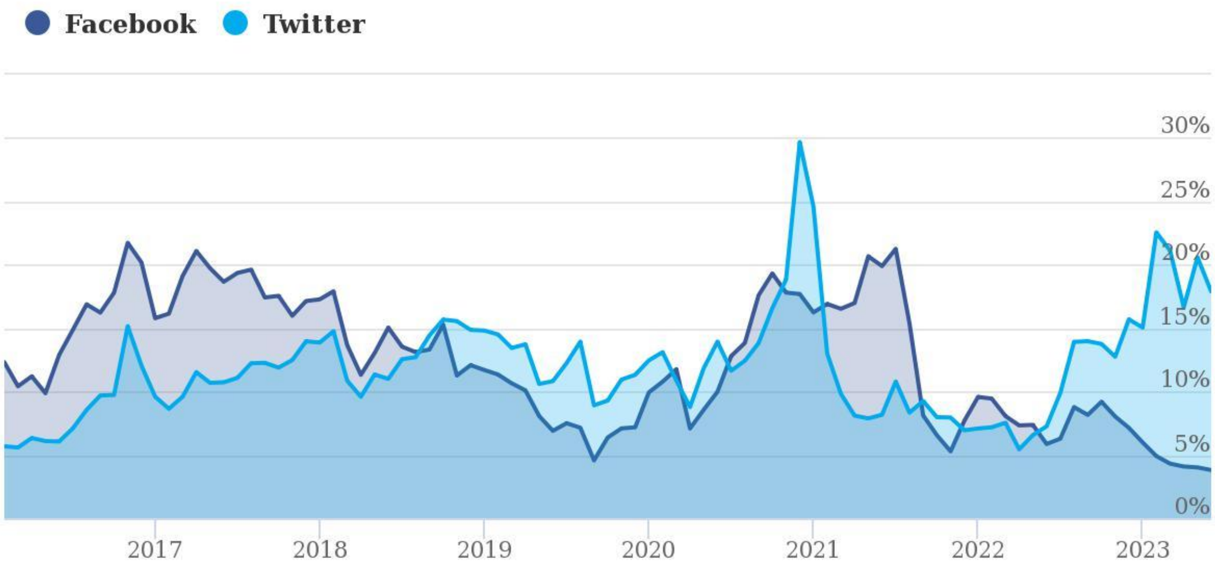


Figure 1. The Iffy Quotient for Facebook and Twitter, dating back to 2016, computed on a monthly basis. See the website for an up-to-date, dynamic version where you can hover over points to see data for particular dates.

Figure 1 shows the Iffy Quotient dating back to early 2016, for both Twitter and Facebook. Note that an Iffy Quotient of 10% does not mean that 10% of all user engagement was with stories that contain misinformation. It means that 10% of all user engagement was with stories *from sites* that NewsGuard (or Media Bias/Fact Check) judged to be in a category that we have labeled Iffy. The Iffy Quotient is most useful as a way to make comparisons across time and between platforms, especially focusing on stable trends rather than single dates.

First, notice the temporal trends. For both Twitter and Facebook, there was an increase in attention to Iffy sites in the run-up to the 2016 U.S. elections and again in the run-up to the 2020 elections. The Iffy Quotient increased on each site from April to November, peaking on or after election day. On Facebook, there was a clear downward trend from about March 2017, reaching a low of 4.6% in August 2019. It again topped 20% in the summer of 2021 but has been below 5% for most of 2023. On Twitter, the Iffy Quotient has been even more volatile, reaching a high of nearly 30% in November 2020. It reached its all-time low in March 2022, at 5.4%, but has climbed dramatically since then, reaching 22.6% in January 2023.

There are several factors that plausibly contribute to these temporal trends. First, in the run-up to the 2016 elections and then afterwards, when people were unusually politically activated, the general public had more interest in political news—especially sensational political news—than in other time periods. URLs from Iffy sites may have been able to appeal to that audience interest better than URLs from other sites. This explanation is consistent with reports of Macedonian sites with no political agenda earning advertising revenue by posting invented or copied political stories.[16] Second, both domestic and foreign publishers with political agendas may have expended more time and money to spread misinformation during the run-up to the election. We note, however, that there were no similar increases during the 2018 midterm elections.

After the 2016 election, in addition to demand and supply declining, the platforms took some actions to reduce the spread of Iffy content. For example, in December 2016 Facebook announced a partnership with third-party fact-checkers, sending them questionable stories and showing lower in the feed those that the fact-checkers labeled as false.[17] On January 11, 2018, Facebook announced that it would reduce the reach of all public external content in favor of native posts from friends and family.[18] On its own, that wouldn't affect the Iffy Quotient, which is based on whatever public content is most popular. However, that announcement, and one the following week, signaled other changes that might have affected the Iffy Quotient.[19] One change was to prioritize content around which people interacted with friends; it could be that people interact less around content from Iffy sites. Another was to prioritize news that the community rates as trustworthy, that people find informative, and that is local. Without knowledge of exactly when particular product features or policy changes were rolled out, we are not able to assess the impacts of particular initiatives on the Iffy Quotient. Yet there was a long-term decline in Facebook's Iffy Quotient from March 2017 through July of 2019. On August 6, 2018, Facebook announced a ban on several pages associated with InfoWars host Alex Jones. That did not have an immediate impact on the Iffy Quotient, which hovered around 12% in the weeks before and after. Twitter waited until September 6 to take similar steps; similarly, there was not an immediate, measurable effect on Twitter's Iffy Quotient.

Next, notice the difference between Twitter and Facebook. From 2016 through the end of 2018, more Iffy sites gained attention on Facebook than on Twitter. This picture is consonant with the finding of Allcott, Gentzkow, and Yu.[20] They used another commercial service, BuzzSumo, to estimate the total monthly tweets and Facebook engagements for a set of 570 sites that "have been identified as producers of false stories," analogous to our notion of Iffy sites. They found that total Facebook engagements dropped by nearly two-thirds from the end of 2016 to July 2018, while the tweets about URLs from those Iffy sites increased slightly. One advantage of our approach is that by expressing the attention share of Iffy sites as a fraction of that for all popular sites on each platform, we are able to compare the Iffy Quotient for the two platforms directly at any point in time.

By early 2019, the two sites had equalized, and for most of 2019 Twitter had a higher fraction of engagement with URLs coming from Iffy sites than Facebook did. It is not entirely clear why the

[16] https://www.wired.com/2017/02/veles-macedonia-fake-news/
[17] https://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/
[18] https://newsroom.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/
[19] https://newsroom.fb.com/news/2018/01/trusted-sources/
[20] http://web.stanford.edu/~gentzkow/research/fake-news-trends.pdf

two platforms changed positions. Presumably, however, there should have been similar fluctuations in supply and demand for Iffy content at the two sites, with the major difference being policies and technical features. Facebook may have been more successful at detecting and countering fake accounts and manipulation campaigns, more aggressive in discounting ranking signals that are associated with Iffy sites, or more aggressive in demoting particular articles and sources. For much of 2021, Facebook had a higher Iffy Quotient than Twitter, and the sites reversed positions again in mid-2022; in mid-2023, Twitter's Iffy Quotient was more than triple that of Facebook.

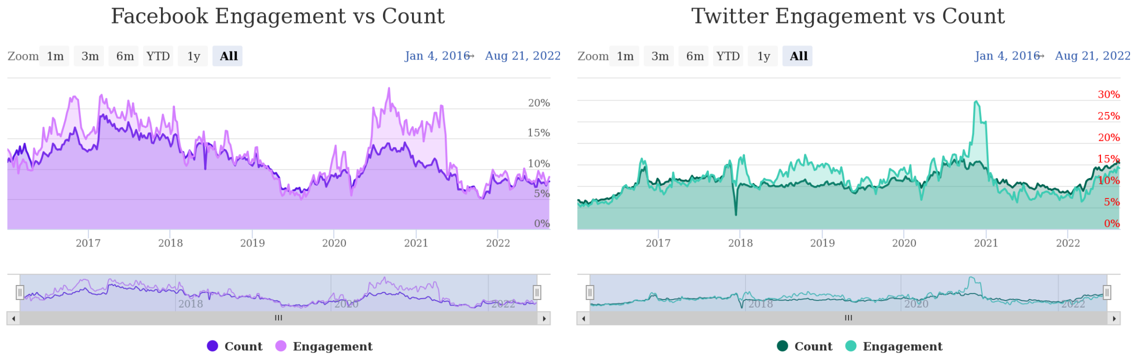


Figure 2. Engagement-weighted vs. count-based Iffy Quotients. On the left, the (mostly higher) lighter lavender line is the same as the corresponding line in Figure 1, the engagement-weighted Iffy Quotient for Facebook, while the darker violet line shows the count-based Iffy Quotient. On the right side, for Twitter, the lighter green line matches that in Figure 1, the engagement-weighted version.

Figure 2 compares the engagement-weighted and count-based Iffy Quotients for Facebook and Twitter. The engagement-weighted version became the primary measure with version 3. For Facebook, in 2016-17 and 2020-21 the engagement-weighted Iffy Quotient was noticeably higher than the count-based. This suggests that URLs from Iffy sites were getting more than the average engagement among the 5,000 most popular URLs.

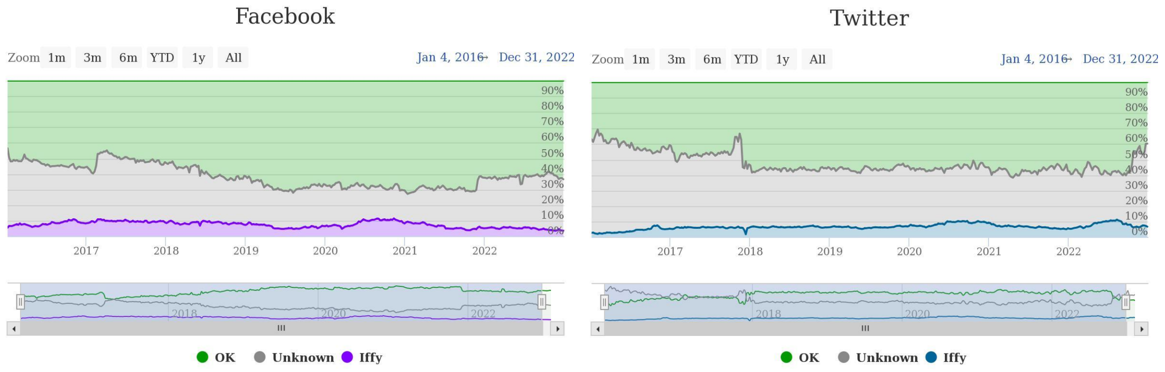


Figure 3: The share of OK and Unknown sites, as well as Iffy sites.

As described in the previous section, an observed decline in the Iffy Quotient could reflect the URL lists becoming stale. Figure 3 shows that the coverage increased substantially from 2016-2022, with a larger fraction of popular URLs coming from sites rated by NewsGuard or Media Bias/Fact Check. Note that NewsGuard started operations in 2018. In late 2021, there was a jump in the fraction of URLs from unrated sites on Facebook; this did not, however, lead to a decline in Facebook's Iffy Quotient, which actually increased during this period. The fraction of popular URLs on Twitter from unrated sites on Twitter grew substantially from mid-October 2022 to the end of the year, from 33% to more than 50%; Twitter's Iffy Quotient also rose during this period.

Tracking Changes in the Iffy Quotient

Beyond the basic trends we have identified above, how should readers make use of the Iffy Quotient? One way is to try to track the impact of external events and of platform technology and policy changes. This can be done retrospectively, as we did for the 2016 elections and Facebook and Twitter's sanctions against Alex Jones.

Tracking can also be done prospectively. When Twitter or Facebook announce a counter-measure, such as deletion of a large number of fake accounts, journalists can start to track the Iffy Quotient and see if it changes.

Or you can sign up for alerts. Join our email list using the form at our website: https://csmr.umich.edu/contact/. We will send out an alert when there is a significant change in the Iffy Quotient for either Facebook or Twitter that is sustained over several days, as well as other general announcements from the Center for Social Media Responsibility, such as the release of additional Platform Health Metrics.

What's Next for the Iffy Quotient?

Our website that tracks the Iffy Quotient will automatically update daily. We will be making iterative improvements, and we welcome partnerships with organizations that can help us validate or improve the metric. In the URL collection phase, we would welcome other sources that could be combined with what we get from NewsWhip. In the classification phase, we would like to reduce the large number of Unknown sites. We are pleased that the move to NewsGuard in the second version reduced the fraction of Unknown sites and hope that it will be reduced further over time.

We would also like to be able to filter out popular URLs that have nothing to do with news, politics, public affairs, science, or health. That would make the absolute value of the Iffy Quotient a more meaningful quantity—the fraction of popular URLs from Iffy sites among those where reliability of the information matters. It would be especially useful to have an automated classifier that operated on individual URLs rather than entire sites. We would be happy to partner with anyone who has trained a news and public affairs classifier.

We would also like to expand to other platforms beyond Twitter and Facebook. These could include Google search results and YouTube search results and recommendations. If you have data on another platform, or even just a suggestion of how we might collect it, we'd be happy to hear from you.

Conclusion

Platform Health Metrics are a way to provide constructive accountability to the media platforms at a meaningful scale. The current environment, based on identifying and reporting individual bad outcomes, is a less constructive form of accountability for the platforms because they can never show that they are doing well, only making it harder for watchdogs to catch mistakes.

By contrast, the Iffy Quotient tracks trends over time rather than reporting on individual problems. It provides quantitative evidence in support of the claim that Facebook and Twitter did a poor job during the 2016 election season; they amplified the distribution of information from Iffy sites at double the rate that they did earlier that year. But the Iffy Quotient can also tell a more positive story of progress when progress is made, as happened from late 2017 through mid-2019, especially on Facebook.

Acknowledgements

We gratefully acknowledge the NewsWhip service and the work of NewsGuard and Media Bias/Fact Check to assess news sites and share the assessments publicly.

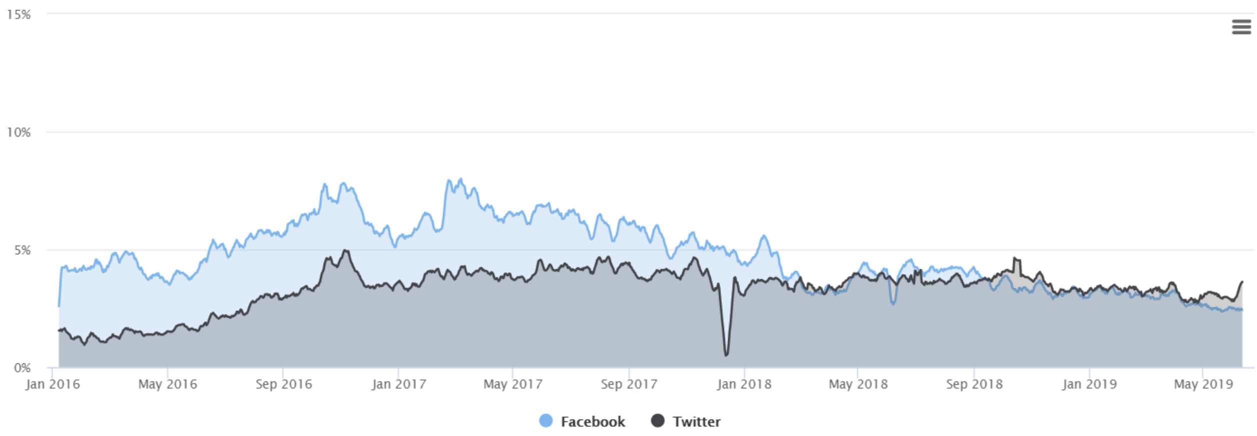
The Iffy Quotient is an adaptation of Aviv Ovadya's previous work toward a dashboard for measuring attention on unreliable sources, which he developed prior to joining the Center for Social Media Responsibility. Ovadya's work began in late 2016 and was also funded through his 2017 Knight News Innovation Fellowship at the Tow Center for Digital Journalism at Columbia University. The Iffy Quotient was also inspired in part by Craig Silverman's analysis of engagements for the top stories from hoax and mainstream news sites.[21]

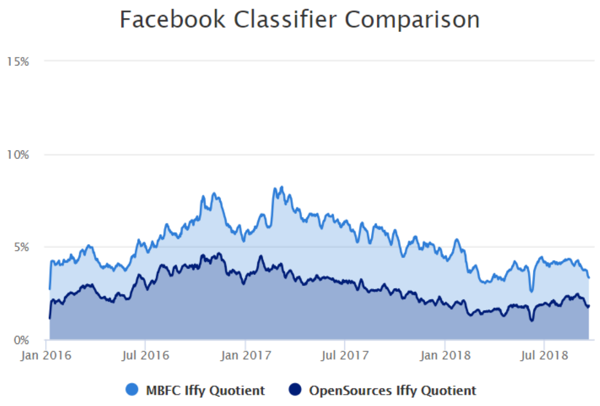
Vitaliy Lyapota was the primary software developer for the Iffy Quotient. Yuncheng Shen created images and animations describing the steps in the calculation of the Iffy Quotient, and Summer Nguyen helped to update them for the second version. Fernand Pajot and Ashwin Rajadesingan conducted data analyses that contributed to quality assurance. James Park contributed editing assistance to the second version of this report. Chenxin Han redesigned the visual display of the Iffy Quotient for version 3 and the new visuals on the website. Siqi Wu contributed to the analytics around deduplication of URLs for version 3. James Park helped to redesign the website and helped edit version 3 of this report.

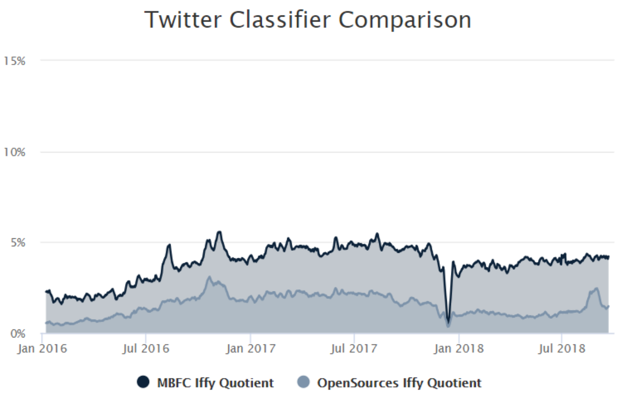
[21] https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook

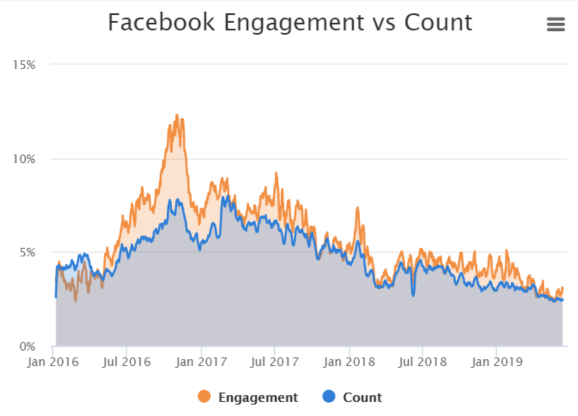
Appendix: Archival Charts From First Version

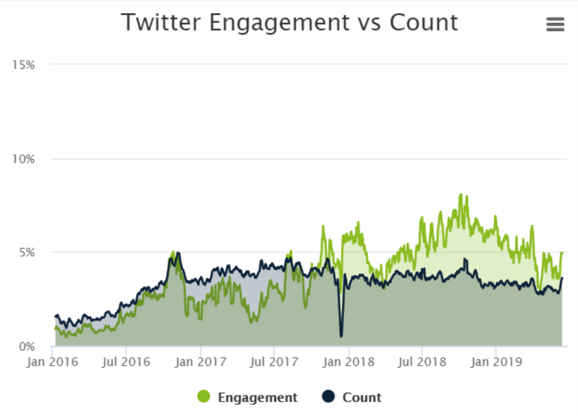
We include archival charts for the first version of the Iffy Quotient, using the December 6, 2018 version of Media Bias/Fact Check ratings, the last that we used on our live site.

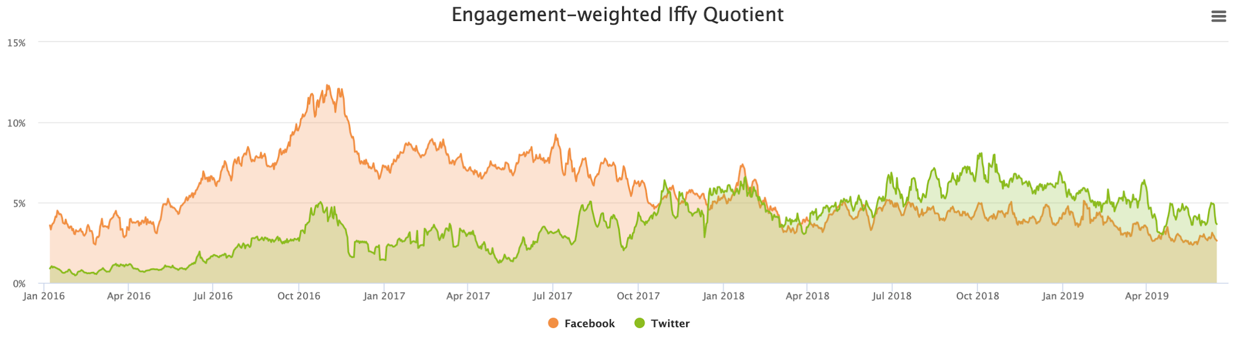


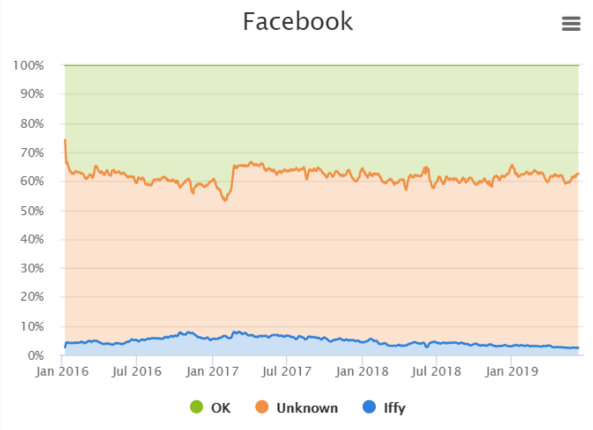


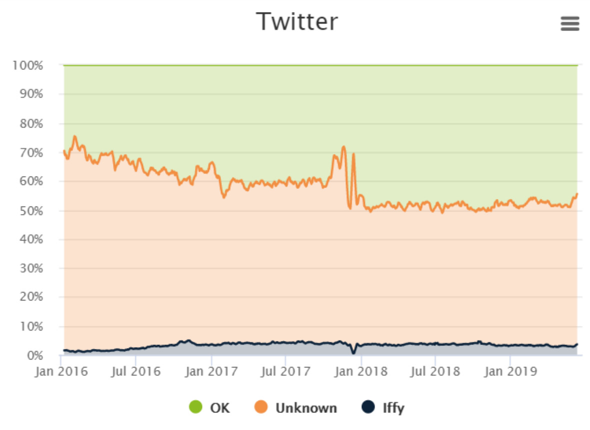







Engagement-weighted Iffy Quotient


Facebook


Twitter