How To:
Verify Online Census Media

Verification is an important process to integrate into your responses to disinformation. While there is no such thing as a foolproof method of verification in the digital age, understanding how to apply certain techniques to verify what you're seeing online is a vital skill—one that can help familiarize you with a variety of disinformation tactics and bolster your public responses. Of course, the threat classification model should be the centerpiece of any response; though it can be tempting to show off the fruits of your verification efforts, your public responses should still hinge on how a threat is classified, so as to avoid unnecessary amplification. And regardless of whether or not problematic content can be verified, it's still always a good practice to flag it to the platforms where it's circulating, as those platforms have a clearer picture of where that content came from, how it's spreading, and whether it violates their policies.

There are four categories of verification techniques you can use:

1. ■ ASSESS PROVENANCE

For any piece of problematic content, you should determine its original version and context to the best of your ability. Are you looking at the original piece of content, or has it been manipulated or repurposed? As First Draft News writes in its "Essential Guide to Verifying Online Information," this is the most important verification process and should always be done first.

- **Census scenarios**

  - A photo is circulating purporting to show an ICE agent examining census data, implying it has been illegally shared outside of the Bureau.
  - A photo or video is circulating purporting to show law enforcement rounding up immigrants after they've filled out the census.

Verification action

Run a reverse image search, which can find identical copies of an image, organized by date and site. You can find many exact matches and thematically similar images. If you find the image was published years earlier, for instance, you'll know it has nothing to do with the 2020 Census, but has instead been repurposed and recontextualized to deceive people. Similarly, if you find an older image that looks almost identical but has something changed—such as the census data on an ICE agent's computer screen, for example—you'll know it was manipulated.

Tools

Google reverse image search: If using Chrome, simply right click on an image and select "search Google for image." Or you can go to this link and either upload an image or paste an image URL by clicking the camera icon. The results will show you links to sites Google thinks are most relevant, visually similar images, and pages that include matching images, where you should pay special attention to the date when a page was published.

TinEye: Go to this link and upload, paste, or enter an image URL. You can sort the results from oldest to newest, which is extremely helpful for verification purposes. Note that TinEye only supports the following file formats: JPEG, PNG, GIF, BMP, TIFF or WebP.

Yandex: Go to this link click the camera icon, to upload an image or paste an image URL. Results will show you similar images and sites where the image is displayed.

InVID verification plugin: (Note: InVID can help perform reverse searches for stills in video content.) Go to this link and download the free plugin, which works with Chrome or Firefox. Once installed, you can click on the inVID icon on your browser and select "Open inVID." There are many tabs and tutorials to explore, and Amnesty International's Citizen Evidence Lab does a good job breaking them all down here. But a good place to start is to click on "analysis" and paste a YouTube, Facebook, or Twitter URL. Once you hit submit, the tool gives you useful metadata associated with the video, such as the upload time and number of likes and shares. It also breaks down social media video into thumbnails which you can then run a reverse image search on, using the tools at the bottom of the analysis page.

RevEye Reverse Image Search browser extension: (Note: This is a great tool because it is a one-stop-shop that aggregates all the other reverse image tools.) Go to this link if using Chrome or this link if using Firefox, and download the browser extension. Once downloaded, simply right click on an image and scroll down to "reverse image search" with the eyeball icon. Select "all search engines."

- **Example: RevEye Reverse Image Search**

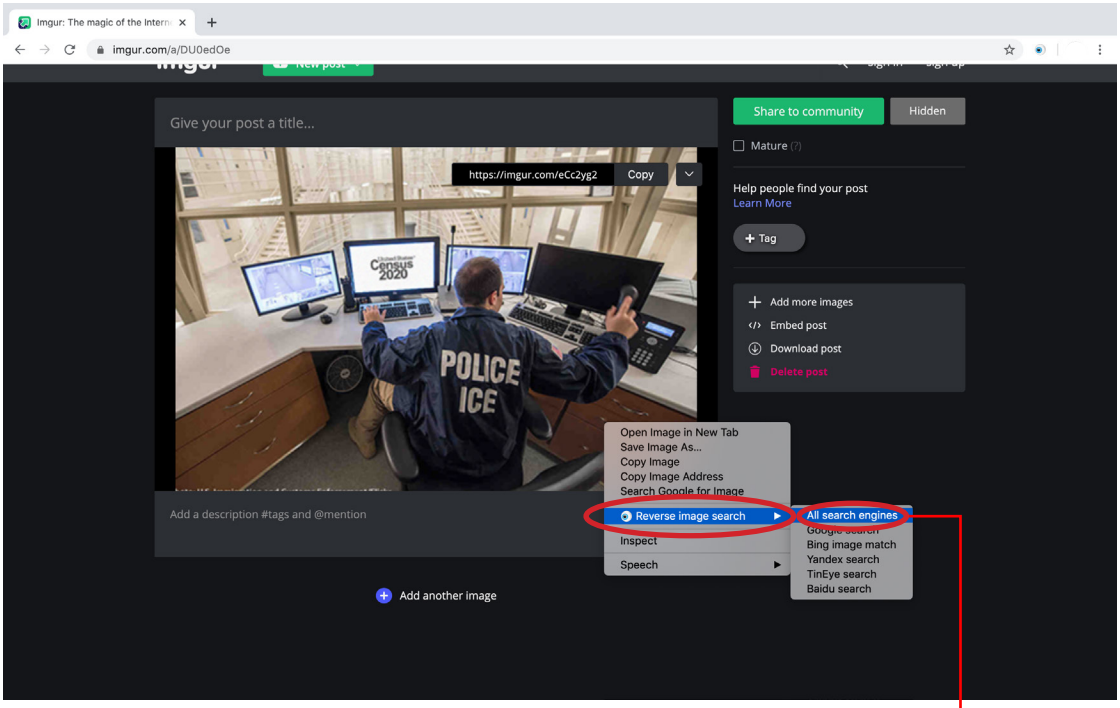


Image 1

Hypothetical example of a problematic image online. (Note: This image was created for the purpose of this demonstration. It was never circulating online.)

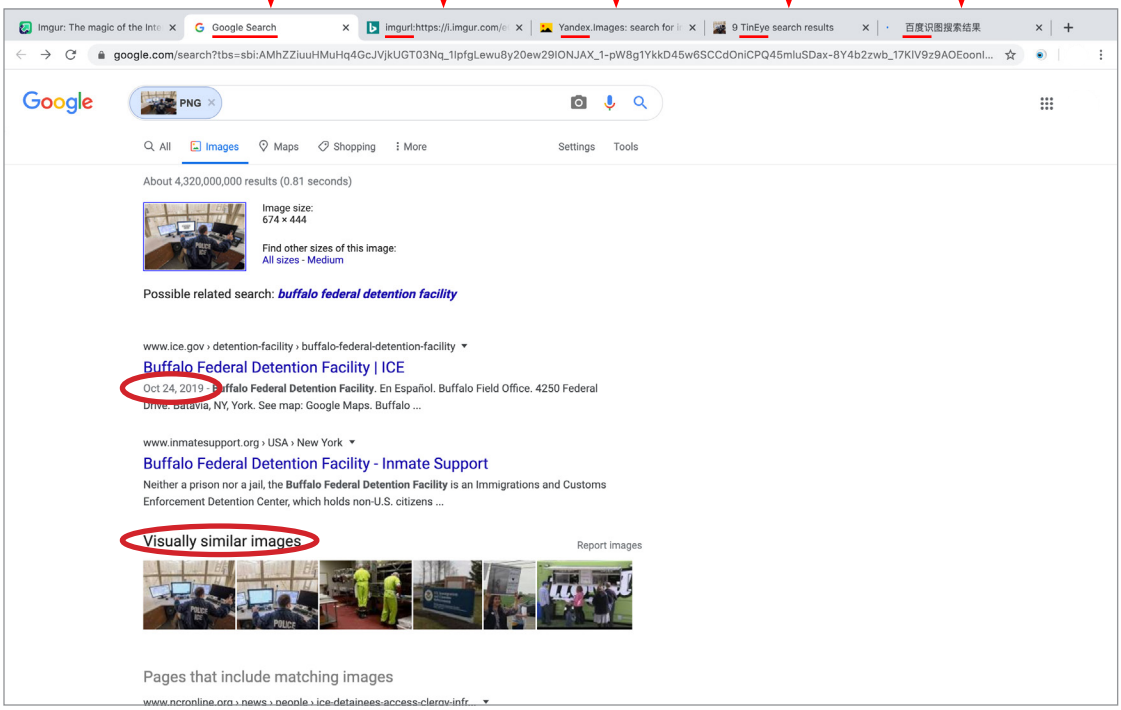
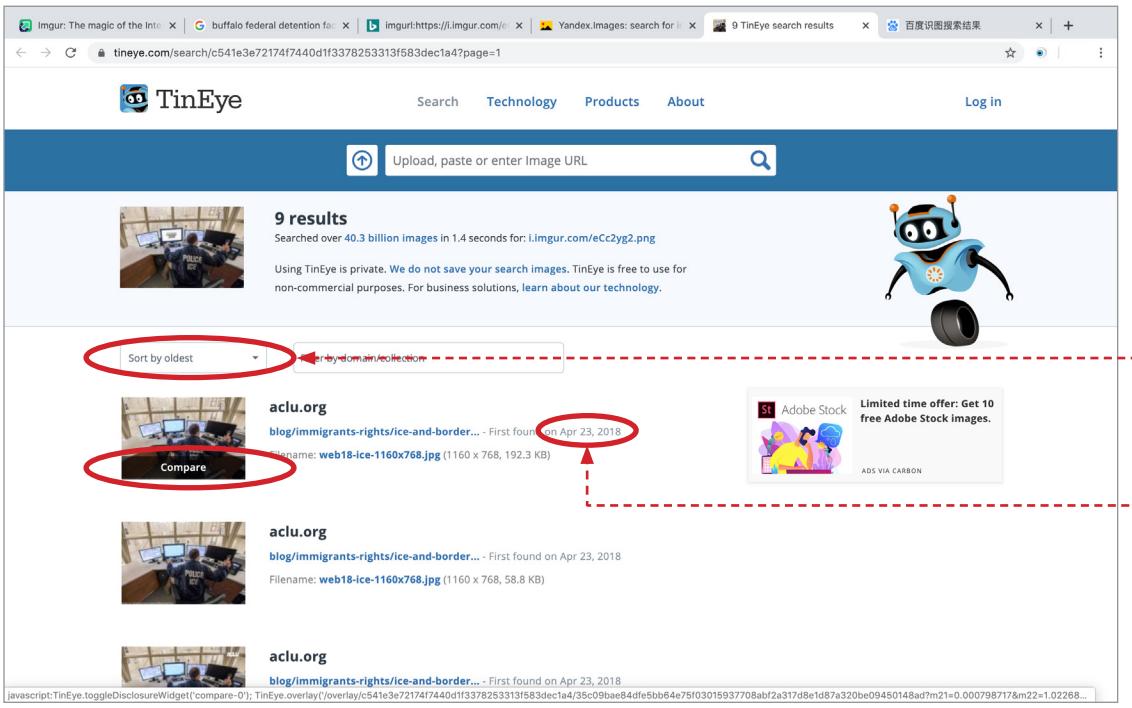


Image 2

The RevEye tool opens a new tab for each search engine. The first new tab shows Google's reverse image search results. Pay special attention to the dates of associated web pages and to the selection of visually similar images.



Change sorting to "Sort by oldest"

Sorting by oldest allows you see the earliest publication date.

Image 3

The second-to-last tab shows TinEye's results. Change to "sort by oldest" to see earliest match. Hover cursor over image and select "compare" to spot manipulated elements.

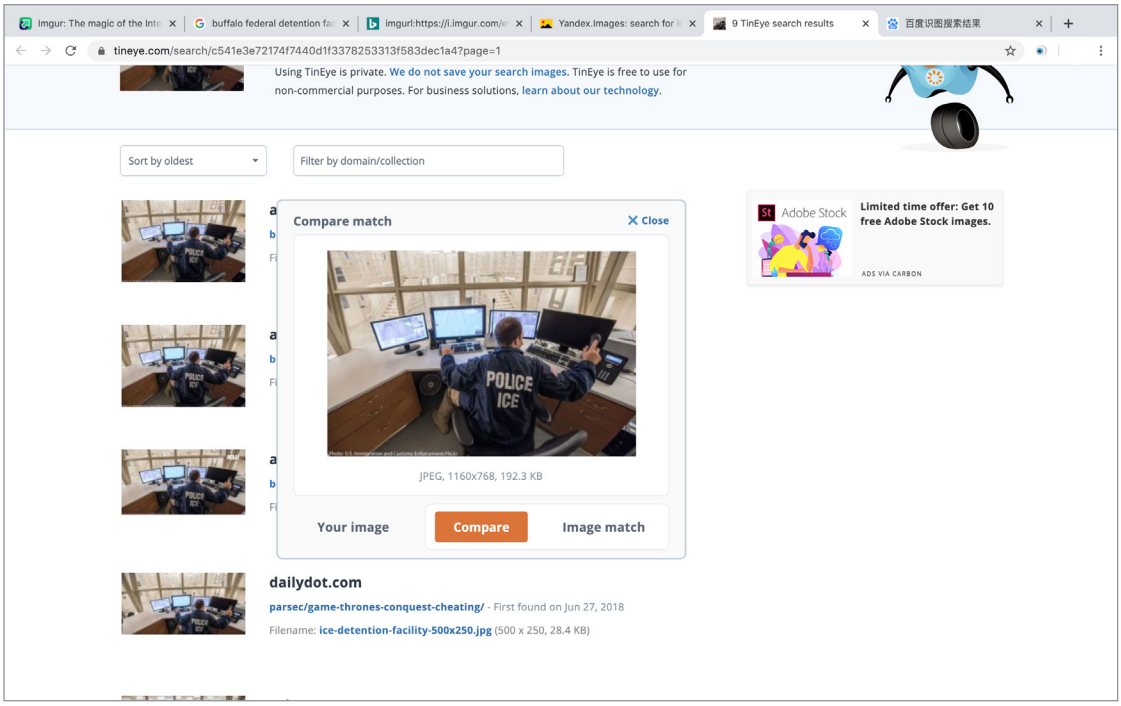


Image 4

By comparing similar images on other websites, you may discover clear evidence that a problematic image has been manipulated.

● Limits to this approach

- Google and Yandex don't order search results by date, so it can be difficult to find the earliest version.

- TinEye and Yandex require direct links to image files, so when searching for content embedded in Twitter posts, for instance, you must first open the actual image file itself in a separate window to get the URL the search engine can use.

- There is no easy reverse image search tool for a full video; you can only run a reverse image search on a screen grab or thumbnail from a video. InVID makes that process easy by breaking down social media video into thumbnails, but sometimes its analysis tool doesn't work if the social media user has enabled certain privacy/sharing restrictions.

2. ■ CONDUCT GEOLOCATION ANALYSIS

Many photos or videos include details that can identify where and when the event occurred. You can use these clues to assess: does the content show an event taking place where and when the account sharing it claims it did?

● Census scenario

A photo or video is circulating claiming to show the enumeration process being misused in a particular place (e.g. ICE raids or home robberies in a specific neighborhood)

● Verification action

Look for visual clues (e.g. signs, license plates, landmarks, weather conditions) that can corroborate the claim that the event took place where and when the account sharing it says it did. For example, if the content shows an event that purportedly took place in a particular neighborhood, can you spot any street signs or monuments that that confirm or contradict this? Or if the content shows an event that purportedly took place in a cold area, can you spot weather conditions or winter clothing?

● Tools

Satellite map imagery: If the photo or video is purportedly from a particular town or neighborhood, type the location into Google Maps. The more specific the information you have about the location, the better. So, if the account sharing the content

says the event in question happened on a particular street and you see a door number in the photo or video, search for that address. From there, you can use the "Nearby" search to check for notable buildings in the source material. For instance, if you see a mosque in a particular video, you can search for "mosque" in the area. The results will give you red arrows where all the mosques in the vicinity are located, narrowing your search. You can also see a satellite image by clicking the box in the lower left-hand corner, and you can browse street view images by dragging the little yellow person icon in the lower-right hand corner into the part of the map you're interested in. The goal is to see if the street view satellite images match up with the images in the photo or video you're analyzing.

Weather reports: Wolfram Alpha is a useful tool that allows you to check the weather in a particular location, on a particular date. So, for example, if you type "weather Denver November 25" into the search bar at the top, you'll see it was snowing that day. Therefore, any photos or videos that claim to show an event taking place in Denver on November 25 should have snow in it.

● Limits to this approach

- Geolocation is difficult if there are no visible structures or other unique identifiers that could connect an image to a place or time.
- Sometimes visual clues can be deceptive. For instance, you might see building signs in a foreign language and be tempted to conclude the image wasn't taken in the U.S. But it might be from a neighborhood with a high concentration of a particular ethnic group, such as Chinatown, Manhattan.

3. ■ INVESTIGATE THE SOURCE

Do you know who shared the content and why?

● Census scenario

A social media user tweets problematic content about the count (e.g. "I'm hearing from my friend who works at the Census Bureau that they're completely unprepared for a cyber attack and that all our personally identifiable information is vulnerable. No way I'm filling out the census and neither should you!")

● Verification action

Analyze social media metadata and search engine results to determine account legitimacy and motivation. Are there any red flags in the metadata that might suggest the

account doesn't belong to the person it claims to belong to, such as a stock photo avatar that's been lifted from somewhere on the internet and is being used for multiple accounts? Can you find any other evidence online that the person behind the account holds particular opinions or biases—e.g. blog posts expressing anti-government views?

Tools

Social media metadata: Follow the below guide for analyzing metadata, as recommended in Data & Society's report "Data Craft: The Manipulation of Social Media Metadata."

Metadata	Red Flags	To Verify
Account Names	Double consonants, default avatars, random numbers; Screen name different from user name; Nonverified account (which is especially suspicious if the account claims to be representing a public official or a high-profile person); Name contains the word 'Official' or 'Real.'	Search for user name and screen name on other platforms.
Banner, Bio, Profile Pics, or Content	Available on Wikipedia, Internet Archive, or other public platforms; Posted to other social media accounts; Far more reposts than 'original' content; Lots of duplicate content.	Search for images with reverse image search tools; Search distinct phrases with "quotes" to discover sources of copied content; Compare to profiles from other platforms with similar account names.
Followers	Recent account creation; Followed by large number of suspicious accounts; Nonsense comments from followers; Sudden growth in followers or following.	Examine the Wayback Machine Archive of account to determine the rate of growth; Conduct steps 1 and 2 with a sample of followers to ensure authentic behavior with account engagement.
Authentic Interaction	Unrelated hashtags; Automated responses from other accounts; Low rates of being liked or shared by other accounts; Content promoted by ad purchase.	Search for conversations, interactions, and activity between the account and followers; Confirm that replies are not simply automated messages, reshares, or responses with links.

Search engines: Search for the account's screen name and user name on Google, Bing, and YouTube and see if the results shed light on the user's possible motivation for sharing the content. Many people use the same username across different sites, so you may be able to find blog posts, comments, or other information related to employment or membership in an organization that can help link the user to a particular viewpoint.

Twitonomy: If the problematic content is posted to a Twitter account, you can sign into Twitonomy and gather useful information about other Twitter accounts, such as when the account joined, its tweet history over a particular range, how many tweets per day it averages, what percentage of its tweets are retweets, what hashtags it uses most, and what users it retweets the most.

● Limits to this approach

- Even when you analyze relevant metadata, illegitimate accounts (such as those that impersonate someone else) can still appear legitimate. Disinformation agents use sock puppet accounts, for instance, to evade platform bans and sway online conversations. These puppet accounts can build up a network of content and followers over time, making it easy to attribute the account to a real person—especially if there are no other accounts tied to that person which you can compare side-by-side.

- Suspicious-looking metadata can also make legitimate accounts appear illegitimate. For instance, you may be convinced that an account is automated because it posts so frequently and never takes breaks, like a human would (e.g. to sleep). Platforms like Twitter, however, make it easy to schedule posts during the day, so an account that posts all the time may still be a real person. *To protect yourself from online harassment or reputational harm, you should never publicly accuse an online account of being a bot.*

4. ■ CHECK FOR AUTOMATED OR COORDINATED INAUTHENTIC BEHAVIOR:

Could the account or content you're seeing be part of a network of bot, human, or cyborg accounts working in tandem to spread disinformation, sow fear and, ultimately, suppress census participation?

● Census scenario

A problematic hashtag (e.g. #boycottthecensus) starts trending on one or multiple social media platforms.

● Verification action

Analyze social media metadata of the accounts heavily using or retweeting the hashtag.

● Tools

Social media metadata: Follow the same guide for analyzing metadata as above, but do so for multiple accounts using the hashtag in question. Pay special attention to whether the accounts are using nearly the exact same language or posting at almost identical times, both of which are red flags.

Bot detection tools: Hoaxy allows you to type in a phrase like "boycott census" and see "bot scores" for the accounts that are using it. Similarly, BotSentinel allows you to analyze an account by clicking the green box on the upper right hand corner. You can enter either a Twitter handle or a tweet URL and see a "trollbot rating," which the site uses to describe "human controlled accounts who exhibit toxic troll-like behavior."

- ### Limits to this approach

  Bot detection tools can only provide a probabilistic estimation of whether an account is a bot based on different behavioral criteria. Results should, therefore, be analyzed in tandem with other metadata.

While there are limits to each of the actions described above, verification is still an important process for anyone operating in today's information ecosystem to understand. It might not be a silver bullet, but when it comes to disinformation, there really is no quick fix. The best we can do is take a number of steps that together will minimize harm, rather than definitively solve the problem. To that end, **it's worth taking some time in the coming weeks to identify the people in your organizations who might carry out this verification work**, much in the same way you've already designated the people who talk to the press or create digital content. And if you have the resources to spare, you might consider investing in Hunch.ly, which is a great tool for documenting online research—something that can be very useful, as digital content sometimes disappears without warning. For more information about these and other verification techniques, check out Amelia Acker's report on "Data Craft," First Draft News's "Essential Guide to Verifying Online Information," and bellingcat's "Beginner's Guide to Geolocating Videos."