

MODULE FOUR:

Navigating Information Disorder

Module Objective: Navigate information disorder to reduce its impact on digital communities

Module Dilemma: I want to flag certain content to the platform I'm moderating my group on and be safe doing it (e.g., remain anonymous)



Fact checking



An important thing all digital community stewards should ask themselves while trying to fact-check a claim is whether it is verifiable, or what is referred to as “fact-checkable”, or not.

To understand this, let’s look at the following examples.

Opinion - An opinion is based on a belief or point of view. It is not based on evidence that can be checked and often displays a bias as it is based on someone’s own experiences, world views and belief systems. (E.g. “Pink is a better colour for girls than blue”).

Prediction - A prediction is a statement predicting or promising something that will happen in the future. E.g., “It will snow in March.”

Fact - A fact is backed up with evidence and can be checked or verified. E.g., “Spiders are arachnids, not insects.”

Thinking something is true because we agree with it, does not make it a fact.

Two people might be in the same room – one saying it is hot and the other that it is cold. These opinions cannot be fact-checked. What can be fact-checked, however, is the exact temperature in the room.

Digital community stewards should only fact-check facts – not opinions or predictions.

It can be a difficult task to figure out what posts and what claims need to be fact-checked. This filtering and prioritisation depend entirely on the admin of the group – who sometimes might be guided and supported by the group members.

In this case, it is important for digital community stewards to focus on the following. Asking these questions can help you decide whether a post should be fact-checked or not.

- Is the topic important?
- Is the claim framed as a statement of fact or opinion?
- Does the claim matter? What is the potential harm?
- Have I flagged content by this group member before?

LET'S TALK ABOUT BOTS

Bots are basically social media accounts that are operated entirely by computer programs and are designed to generate posts and/or engage with content on a particular platform. In disinformation campaigns, bots can be used to draw attention to misleading narratives, to hijack platforms' trending lists and to create the illusion of public discussion and support. Researchers and technologists take different approaches to identify bots, using algorithms or simpler rules based on a number of posts per day.

Below is an infographic from FirstDraftNews, sharing some indicators to "Spot a Bot".

HOW TO SPOT A BOT

Just because it acts like a doesn't mean it is a bot. These indicators of automated or co-ordinated online activity can help, but look for a combination of signs, not just one.

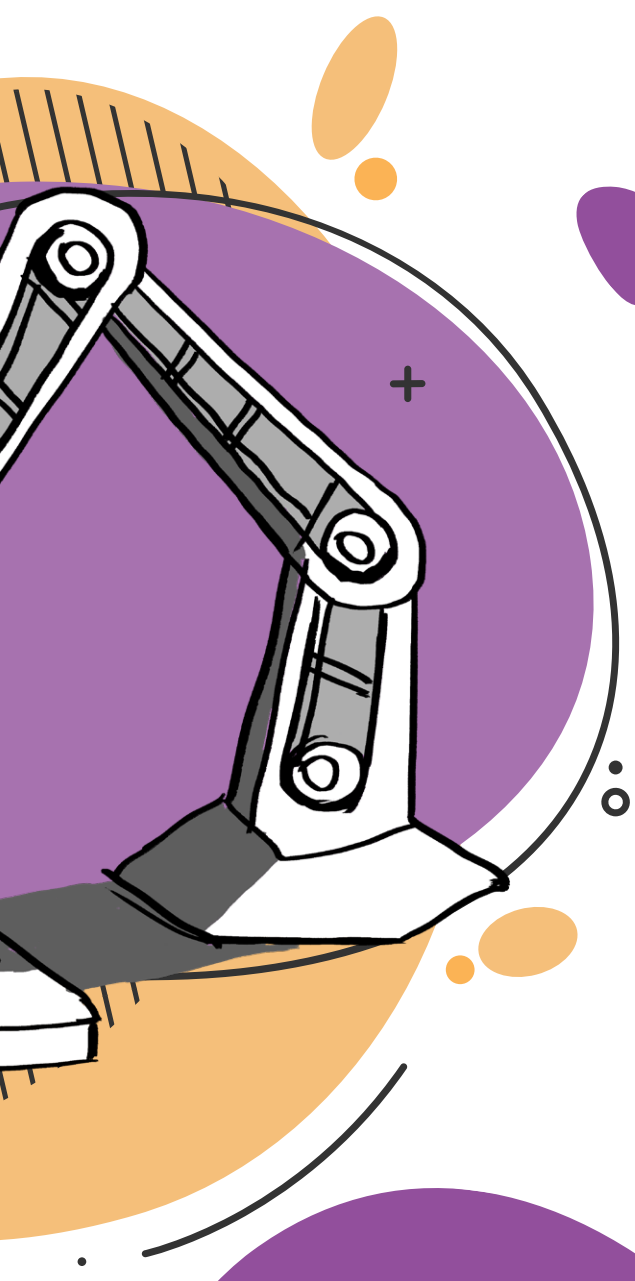
ACCOUNT

- Recent creation date
- Lack of personal information
- Profile photo is ambiguous, stolen or non-existent
- Divisive words, hashtags, URLs or emojis in bio
- Suspicious handle e.g. lots of numbers

CONTENT

- Tweeting in more than one language
- Engaging in multiple international narratives
- Signs of automation or account management software like buff.ly
- Posting inflammatory memes or GIFs
- Hashtag spamming
- Occasional off-brand retweets
- Very few reliable news sources
- Awkward turns of phrase





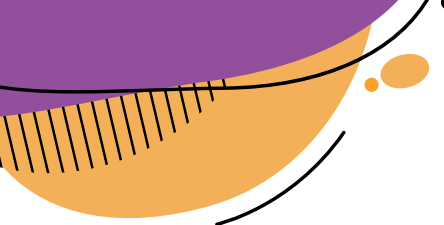
NETWORK

- Followers and following is high and almost identical
- High number of following and no followers
- Following a suspicious mix of sources
- Connected to other suspicious accounts
- Duplicated account
- Previously circulating suspicious content
- Previously identified by other organisations

More information on the above indicators can be found [here](#).

ACTIVITY

- High volume of tweets (more than 100/day)
- High percentage of retweets (more than 80s%)
- Posting persistently day and night
- Posting only at specific times of day
- Sudden spike in activity or change in interests



Several infographics have been developed, explaining to people how to fact-check fake news, mis and disinformation.

We have adapted an infographic from the International Federation of Library Associations ([IFLA](#)) and [Simon Fraser University](#) to share some general and key guidelines for fact-checking:

GUIDELINES FOR FACT CHECKING

Consider the source:

Click away from the story to find the site and contact info

Read beyond:

headlines can be developed to be outrageous, creating “click bait” for more sharing. Read the entire article.

Check the author:

Quickly search the author, and ask: Are they credible or controversial? Are they real?

Supporting sources:

Click on the sources in the article, and ask: Does the information accurately support the story?



Check the date:

Is this old news? Reposting old news may not mean it is relevant to more current events.

Is it a joke?

If it is extremely difficult to believe, it might be satire. Look into the site and article to be sure.

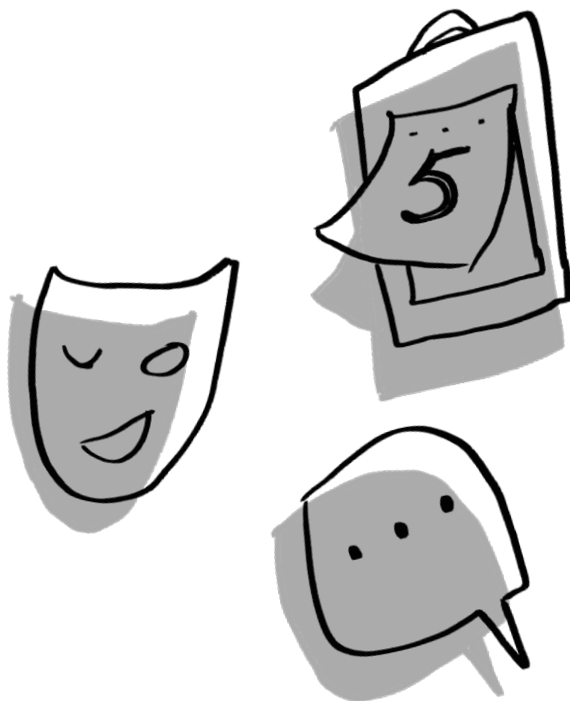
Check your biases:

Could your own position on the subject or beliefs be affecting your judgement? Remember to take a step back and think critically about this before making judgements.

Misinformation and Disinformation is created to appeal to our emotions and instincts. Think about why this news is making you react in a certain way.

Ask the experts:

Ask a topic expert, or cross-check it with other trusted or official sources or a fact-checking site.



You can watch [this video](#) to learn more about these key factors of spotting mis and disinformation

An interesting resource: [why do we prefer information disorder over fake news?](#)

Verification

WRITTEN INFORMATION

BOOLEAN SEARCH QUERY

Smart searches cut through social media chatter by finding precise snippets of information based on keywords. When searching for newsworthy content online, you've got to know exactly what you're looking for and have the skills to find it. Using the right keywords to search in the right places is key. This is where Boolean search queries help. These strings of words allow you to cut through the usual social media chatter by upgrading a default search to a multifaceted, specific search to find more precise snippets of information.

Boolean search is used when you require stricter search criteria to eliminate generic results that are loosely connected to your brand or other keywords.

Boolean searches help you to specify exactly what you are looking or not looking for. For example, let's say you're searching for posts during a breaking news event, such as the Notre Dame fire. You want to search for Notre Dame, but you won't want posts about the Disney film. A boolean search will allow you to include posts that mention "Notre Dame" but exclude ones about the Disney film to refine your search results and find the information you're after.

This is possible with 'operators', which allow you to combine multiple keywords. There are three operators for basic searches: AND, OR, and NOT.

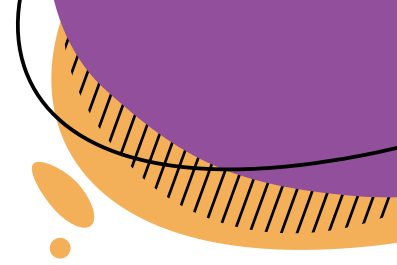
A quick note: Twitter, Reddit and Google currently support Boolean searches, while YouTube supports them to some extent on top of its advanced search function. Facebook used to support Boolean searches, though it appears it no longer does.



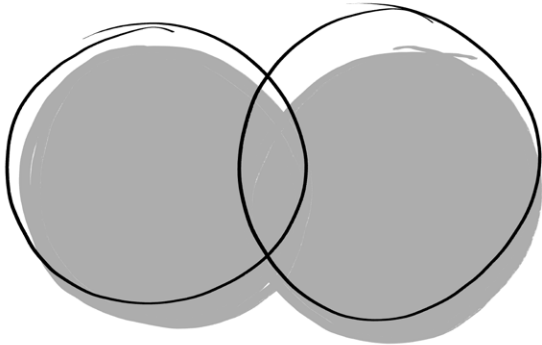
AND – This allows you to narrow your search to only retrieve results that combine two or more terms. For example, you might want to search for "Notre Dame" and fire.

OR – This allows you to broaden your search to retrieve results connecting two or more similar terms. This can be good for misspellings and typos. In the case of Notre Dame, you could search for "Notre Dame" OR "Notre Dam". This will retrieve all results containing either phrase.

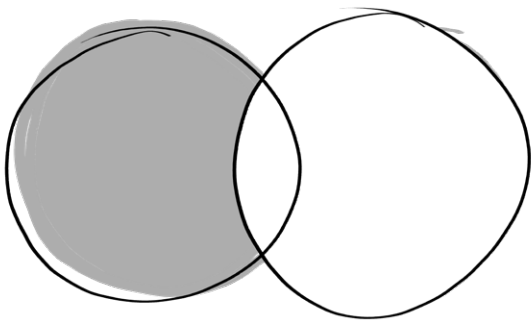
NOT – This allows you to exclude terms you don't want to appear in your search results. For example, you might want to exclude mentions of the Disney film The Hunchback of Notre Dame, and so you could write "Notre Dame" NOT "hunchback".



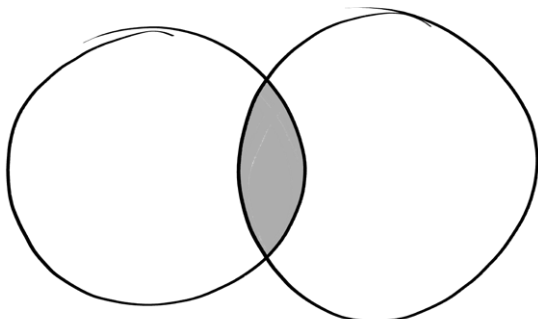
“Notre Dame” OR fire



“Notre Dame” NOT fire



“Notre Dame” AND fire



Key points:

- Operators (AND, OR, NOT) must be written in capitals, or they won't work
- If you're searching for phrases (terms made up of multiple words) then you have to put them in quotation marks (eg “Notre Dame”)
- You won't be able to find information that has been made private by a user

You can find more ‘operators’ for Boolean searches [here](#).

Traditionally, to search for content within a certain timeframe, you can search using the “since:” and “until:” operators (on Google it's “before:” and “after:”).

For example, if looking for examples of people turned away from a polling booth on the day of the European Parliament elections, you could search: “polling booth” AND “turned away” near:London since:2019-05-23 (year-month-day).

You can also look for content posted up to a certain date. For historical examples of people being turned away from polling booths, search: “polling booth” AND “turned away” until:2019-05-22.

Twitter, Google and YouTube also all provide dates in their advanced search functions which allow you to select a custom date frame.

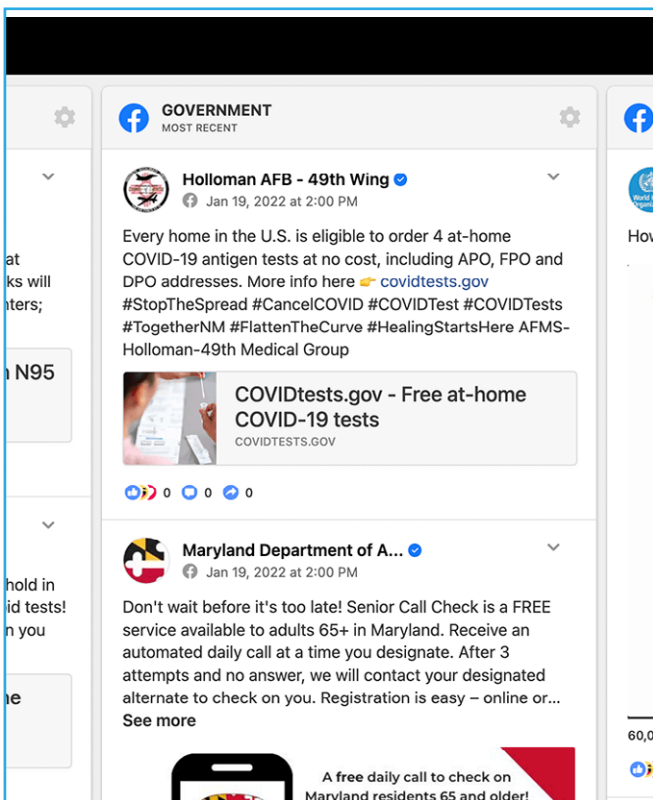
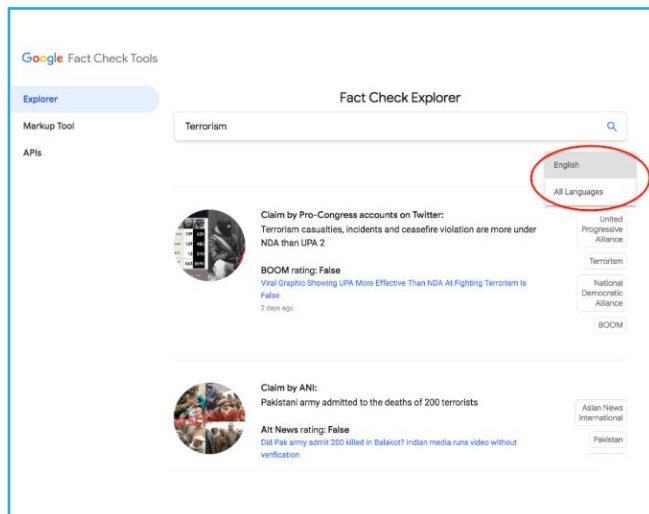
To search for social posts within a geographic area to find a source, you can search for content that was published in a specific location.

While very few social media users geotag their posts, it's possible to search for the specific locations of the ones that do.

For example, if there was a lack of booths at a London polling station, you might search: “polling booth” near:London within:15mi. You could even search: “polling booth” AND “turned away” near:London within:15mi.

FACT CHECK EXPLORER – GOOGLE

This tool allows you to easily browse and search for fact checks. For example, you can search for a politician's statement or for a topic. You can also restrict results to a specific publisher. You can search by keywords and see a list of matching claims and the corresponding fact checks. Keywords can be anything from specific topics to full politician quotes. Basically, anything you would also put into the regular Google Search bar.



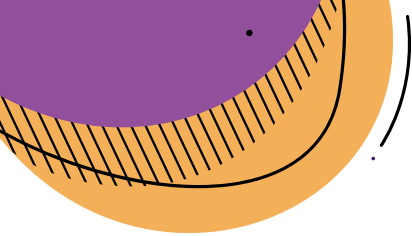
CROWDTANGLE LIVE DISPLAYS ON COVID-19

This is a very specific and [very useful resource](#) to navigate information disorder regarding the COVID-19 pandemic.

CrowdTangle is a social analytics tool owned by Facebook. Their main portals require sign-up, but everyone can access their Public Live Displays. It's a quick, visual way to see how information on coronavirus is being spread on social media.

Public Live Displays are organised by region and country and show content from local media, regional World Health Organization pages, government agencies, and local politicians, as well as social media discussion from Facebook, Instagram, and Reddit.

Each Public Live Display shows Covid-19 related posts in real-time, sorted by keyword, with public pages and accounts for each region.



VISUAL INFORMATION



A picture is worth a thousand words, and when it comes to disinformation it can also be worth a thousand lies. One of the most common types of misinformation we see at First Draft looks like this: genuine photographs or videos that have not been edited at all but get reshared to fit a new narrative. Let's look at such an incident.

In January, Facebook posts receiving thousands of shares featured the photograph (embedded above) and claimed the people in the photo were coronavirus victims in China. A quick look at the architecture shows that it looks very European, which might raise suspicion. Then if we take the image, run it through a reverse image search engine, and look for previous places it has been published, we find the original from 2014. It was an image, originally published by Reuters, of an art project in Frankfurt, which saw people lying in the street in remembrance of the victims of a Nazi concentration camp.

But with a few clicks, we can verify these types of images when they are shared online and in messaging groups.

Just like we can “Google” facts and claims, you can ask a search engine to look for similar photos and even maps on the internet to check if they've been used before. This is called a ‘reverse image search’ and can be done with search engines like Google or Bing.

REVEYE REVERSE IMAGE SEARCH

[This \(chrome\) extension](#) allows performing an inverse image search by right-clicking on any image in a website. You can configure the context menu to contain either a single button with your default search engine or a cascaded menu with all included search engines.

This extension allows us to perform a search by image. It even allows us to choose between the image search engines Google or Bing.

A sample tutorial can be found [here](#).

THUMBNAILS AND VIDEOS

Whenever we upload a video to the internet, it creates a thumbnail or screenshot to show as a preview. We can manually change it, but most people don't. Just like we can use a reverse image search to find out if a photograph has been published on the internet before, we can use thumbnails to see if a video has been previously posted online. Using reverse image search, you can take several thumbnails from any video and check whether it's been posted on the Internet before.

INVID

The InVid project develops a video verification tool called WeVerify to detect emerging stories and assess the reliability of video content spread via social media. It is a plug-in toolkit designed to assist in fact-checking through video verification. The tool provides users with contextual information on videos, reverse image searching, video metadata, and video copyright information, along with other features to assist in verifying content.

A sample tutorial can be found [here](#).

Watch this video, which looks at a case study on verifying a social media post.

Flagging

HOW DO WE FLAG CONTENT?

What comes after is reporting such information to the relevant social media platforms. For this purpose, take a look at [this resource](#) from WHO, and additional resources, to inform yourself about the reporting practices relevant to your platforms. This resource will help you understand how to report and flag content on social media platforms such as: Facebook, Instagram, Twitter, WhatsApp, YouTube, TikTok, Discord, LinkedIn, Viber, etc.

Note: An important aspect of flagging such content in social media groups is to ensure the members do not repeatedly post or spread such content. Therefore, an essential part of this process is to address this issue with the member who posted the misinformation or disinformation in the group.

