

Guide to Monitoring Image and Video-based Social Media



Guide to Monitoring Image and Video-based Social Media

About Democracy Reporting International

Democracy Reporting International (DRI) strengthens democracy by shaping the institutions that make it sustainable. We support local ways of promoting democracy with impartial analysis and good practices, bringing international standards to life. The belief that people are active participants in public life, not subjects of their governments, guides what we do. We work with local actors to protect and expand our shared democratic space in a polarized world, regardless of political opinions or personal beliefs. Find out more at: <http://www.democracy-reporting.org>

Acknowledgements

This guide was written by Maeve Sneddon, with contributions by Madeline Brady. The topics included in this guide were selected based on feedback from 15 researchers and civil society organisations monitoring social media. Forset designed the layout of this publication.

Date: June 2021

This paper is part of a project funded by NEF-Civitates. Its contents do not necessarily represent the position of NEF-Civitates.




 This publication is available under a Creative Commons Attribution Non-Commercial 4.0 International license.

Table of Contents

Glossary

- 1. Introduction 8
- 2. Manipulated Media Content as a Disinformation Threat 9
- 3. Gathering Image and Video Posts 14
 - a. YouTube 15
 - b. Instagram 20
 - c. TikTok 24
- 4. Analysis of Images and Videos 28
 - a. How to Verify Individual Pieces of Content 28
 - b. How to Detect Deepfakes 38
 - c. How To Monitor on a Large Scale 39
- 5. Storing Image and Video Files 42
 - a. Organizing Content 42
 - b. Archiving Content 44
- 6. Exercises 46

Glossary

API application: programming interface; a set of functions and procedures allowing for the creation of software applications; allows developers or researchers to access a software application.¹

Artificial intelligence (AI): the ability of computers to perform tasks associated with human intelligence; projects being developed using this ability.²

Cheapfakes: media products manipulated with a low level of technological sophistication.³

Chronolocation: the process of determining when an image or video was created.

Convolutional neural network: a deep learning algorithm that takes an image as an input, assigns importance to aspects of the image, and then uses that information to differentiate between images.⁴

Deepfakes: media products manipulated with a high level of technological sophistication, often involving AI technologies.⁵

Deep learning: a subfield of machine learning concerned with algorithms inspired by brain function.⁶

Disinformation: false information that is spread purposefully.⁷

Generative adversarial networks (GANs): a deep learning method that allows computers to automatically discover and learn patterns in data, and then to create content that fits those patterns.⁸

Geolocation: the process of determining where an image or video was created.

Hashtag: a label for content, commonly used on social media sites.⁹

Hate Speech: any kind of communication that attacks or uses discriminatory language against a person in relation to their religion, ethnicity, sexual orientation, gender identity, nationality, race, colour, descent, gender or other identity factor.¹⁰

Machine learning: an application of artificial intelligence that allows systems to automatically learn and improve themselves.¹¹

Manipulated media content: any images or video that have been altered or presented in a misleading way.

Misinformation: incorrect or misleading information, not necessarily spread purposefully.¹²

Metadata: data about other data; information about the rights and administration of an image or video; can include identification of the creator, of the model used to create the image or video, of the date and time created, of the date and time posted, etc.¹³

Search term: a word or phrase entered into a search engine.¹⁴

Social media: websites and applications allowing users to share content.¹⁵

Verification: the process of establishing the truth, accuracy or validity of a piece of content.¹⁶

Virtual machine: an isolated, self-contained environment run on a computer program that simulates another computer.¹⁷

1. Jonathan Freeman, “What Is an API? Application Programming Interfaces Explained”, InfoWorld, August 8 2019,

<https://www.infoworld.com/article/3269878/what-is-an-api-application-programming-interfaces-explained.html>

2. B.J. Copeland, “Artificial Intelligence”, Encyclopedia Britannica,

<https://www.britannica.com/technology/artificial-intelligence>

3. Rafael Goldzweig and Madeline Brady, “Deepfakes: How Prepared Are We?”, Democracy Reporting International, November 2020,

<https://democracy-reporting.org/wp-content/uploads/2020/11/2020-11-Deepfakes-Publication-No-2-Web-file-1.pdf>

4. Sumit Saha, “A Comprehensive Guide to Convolutional Neural Networks—the ELI5 Way,” Towards Data Science, December 17 2018,

<https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>.

5. Goldzweig and Brady, “Deepfakes: How Prepared Are We?”, op. cit., note 3.

6. Jason Brownlee, “What Is Deep Learning?”, Machine Learning Mastery, August 16 2019,

<https://machinelearningmastery.com/what-is-deep-learning/>

7. “Definition of Disinformation,” Merriam Webster,

<https://www.merriam-webster.com/dictionary/disinformation>

8. Jason Brownlee, “A Gentle Introduction to Generative Adversarial Networks (GANs),” Machine Learning Mastery, June 16 2019,

<https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/>.

9. Anita Campbell, “What Is a Hashtag? And What Do You Do With Hashtags?,” Small Business Trends, August 12 2013,

<https://smallbiztrends.com/2013/08/what-is-a-hashtag.html>

10. “UN Strategy and Plan of Action on Hate Speech”, United Nations, May 2019,

<https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20Plan%20of%20Action%20on%20Hate%20Speech%2018%20June%20SYNOPSIS.pdf>

11. “What Is Machine Learning? A Definition,” Expert.ai, May 5 2020,

<https://www.expert.ai/blog/machine-learning-definition/>

12. “Definition of Misinformation”, Merriam Webster,

<https://www.merriam-webster.com/dictionary/misinformation>

13. “What Is Metadata and Why Is It as Important as the Data Itself?,” Open Data Soft,

<https://www.opendatasoft.com/blog/2016/08/25/what-is-metadata-and-why-is-it-important-data>.

14. “Search Term: What Are Search Terms?”, Word Stream,

<https://www.wordstream.com/search-term>

15. “Definition of Social Media”, Merriam Webster,

<https://www.merriam-webster.com/dictionary/social+media>

16. “Definition of Verification”, Merriam Webster,

<https://www.merriam-webster.com/dictionary/verification>

17. “What Is a Virtual Machine and How Does It Work”, Microsoft Azure,

<https://azure.microsoft.com/en-us/overview/what-is-a-virtual-machine/>

1. Introduction

Video manipulation tools are increasingly cheaper and easier to access, and the popularity of video-based platforms like TikTok, YouTube and Instagram continues to rise. Image and video posts may be even more powerful than text in impacting human emotions and serving as tools to confuse or manipulate public perceptions of political issues. At the same time, there is a lack of systematic analyses for understanding image and video disinformation patterns or of studies of their effects on visually based platforms.

This guide seeks to provide comprehensive image and video resources for the monitoring and gathering of data from web-based platforms. It explains how to analyse individual pieces of content, provides tips and examples on how to create coding categories for monitoring efforts, suggests tools and techniques for the large-scale monitoring and storage of data, and offers a review of a few current deepfake detection tools available to users. The information provided here will take researchers through the process of studying manipulated media content from start to finish. Below is a roadmap of what readers can expect to gain a better understanding of from this guide.



2. Manipulated Media Content as a Disinformation Threat

What is manipulated media content and why is it a threat?

Just as with text-based materials, image and video content can contain false information. This type of content has been labelled “manipulated media”. For the purpose of this guide, we will use the term “manipulated media content” to refer to any image or video that has been manipulated or presented in a misleading manner. This can include anything from deepfakes, which are images or videos altered using artificial intelligence (AI) or other sophisticated means, to images simply shared with incorrect captions. The creation of deepfakes requires a high level of technological sophistication and cannot usually be carried out by the average person. “Cheapfakes” is a term used to describe all other manipulated media content. These can typically be produced using readily accessible software and involve low levels of technological sophistication.¹⁸

Media manipulation has been used around the world for misinformation and disinformation purposes increasingly in recent years, particularly in the context of elections (see examples in Table 1). Just as with text-based content, manipulated media content and the information it contains can lead to the spread of false information online, growing mistrust in the mainstream media, and the radicalization of online users.¹⁹

At the same time, image- and video-based platforms have grown increasingly popular in recent years. The immense growth of the video-based platform TikTok, which is now the seventh most popular social media platform in terms of users, demonstrates the immense potential of image and video content.²⁰ Now boasting over 600 million users, TikTok is just one platform that shows the reach of image and video content. Recent research shows that image and video content can be both more memorable and more widely shared than text posts.²¹ This contributes to the ability of manipulated media content to go “viral” and spread misinformation to countless users in a short period of time, thus giving image and video content the potential to spread misinformation rapidly.

18. Goldzweig and Brady, “Deepfakes: How Prepared Are We?”, op. cit., note 3..

19. Alice Marwick and Rebecca Lewis, “Media Manipulation and Disinformation Online”, Data & Society, http://www.chinhnghia.com/DataAndSociety_MediaManipulationAndDisinformationOnline.pdf

20. Maryam Mohsin, “10 TikTok Statistics You Need to Know in 2021”, Oberlo, February 16, 2021, <https://www.oberlo.com/blog/tiktok-statistics>

21. Georg Stenberg, “Conceptual and Perceptual Factors in the Picture Superiority Effect,” European Journal of Cognitive Psychology 18 (November 1, 2006): pp. 813–47, <https://doi.org/10.1080/09541440500412361>;
Diana Dilworth, “INFOGRAPHIC: Videos Shared 12x More Than Texts, Photos Liked 2x More Than Text,” InfoWeek, August 29 2012, <https://www.adweek.com/performance-marketing/infographic-videos-shared-12x-more-than-texts-photos-liked-2x-more-than-text/>.

Table 1: Social Media Users per Major Platform







Platform	Number of Users Worldwide in January 2021 ²²
 Facebook	2.7 billion
 YouTube	2.2 billion
 Instagram	1.2 billion
 TikTok	689 million
 Twitter	353 million

 image- and video- based platforms

TikTok is not the only platform by which misinformation can be spread. Image- and video-based platforms like YouTube and Instagram, which have many more users worldwide, are also likely to be sites of manipulated media content. There have also been many instances of misinformation and radicalisation on Facebook and Twitter, which are more traditionally text-based. The reach of these platforms makes it clear that it is important to monitor manipulated media content across all platforms.²³

To provide more clarity on the wide range of manipulated media content that researchers may come across, below are some examples.

Table 2: Examples of Manipulated Media Content

RECONTEXTUALIZING	EXAMPLE
Recontextualizing refers to images or videos that have not necessarily been altered, but that are being presented in a misleading way. This often comes in the form of images or videos posted with incorrect captions about where or when they were taken. Recontextualizing is very common in online political discourse, and recontextualized content is often spread accidentally by average users.	<div><p>The claim:</p><p>This video was shared in October 2019 with the caption “Mexico is in a virtual state of war” at the time of violent clashes in the Mexican city of Culiacan, Sinaloa, between drug cartel gunmen and the military. Click play video includes some profane language.</p><p>Francisco Guillén</p><p>The reality:</p><p>This video is authentic. However, it was not filmed in Culiacan in October 2019. It was filmed in February 2017 by an eyewitness who captured the moment Mexican marines fired from a helicopter during clashes with drug cartel members in a city called Tepic in Nayarit.</p></div>

22. “Most Popular Social Networks Worldwide as of January 2021”, Statista, <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
23. H. Tankovska, “Most Used Social Media 2021,” Statista, February 9 2021, <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>

Reuters discovered the instance of a video shared in October 2019 that claimed to show violent clashes between drug cartels in the Mexican city of Culiacan. While the video itself is authentic and not altered, it was actually filmed in February 2017 in the state of Nayarit. This kind of manipulation is often used to make users believe that events are more violent or drastic than they actually are, as is the case with this recontextualized video claiming there was a drug war in Culiacan.²⁴

MEMES

According to Data & Society’s Deepfakes and Cheapfakes Report, memes can be defined as “images that quickly convey humor or political thought”.²⁵ They are shared widely on social media and are often present in political discussions related to elections, and can contain both images and text. Memes are also often sources of misinformation or hate speech, making them very highly relevant to election monitoring research. Any users can create or share memes, making them a common form of manipulated media content.

EXAMPLE



A NYTimes discussion of the “Pepe the frog” meme, which was recently classified as a hate symbol after its continued use in white supremacist circles.²⁶



SPEEDING, SLOWING, AND CROPPING

Videos can easily be sped up, slowed down or cropped in misleading ways. Speeding, slowing and cropping are commonly used in election contexts to smear candidates. This type of manipulation can be performed by the average user of relatively simple technology and has become a growing source of online political misinformation.

EXAMPLE



24. “Manipulated Media,” Reuters, <https://www.reuters.com/manipulatedmedia/en/chapter-1-manipulated-media#lost-context>
25. Britt Paris and Joan Donovan, “Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence”, Data & Society, https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1.pdf
26. “Can a Meme Be a Hate Symbol?”, The New York Times, October 3 2016, <https://www.nytimes.com/roomfordebate/2016/10/03/can-a-meme-be-a-hate-symbol-6>
27. Paris Martineau, “How an InfoWars Video Became a White House Tweet,” Wired, November 18 2020, <https://www.wired.com/story/infowars-video-white-house-cnn-jim-acosta-tweet/>

PHOTOSHOPPING/VIDEO EDITING	EXAMPLE
Basic editing software, such as Photoshop, can be used to manipulate footage and images and mislead the viewer. These edits are more difficult to make than speeding, slowing or cropping videos, but can still be made by amateur users and be spread easily online.	 <p>A false BBC news video was released claiming the outbreak of nuclear war. The video showed a “BBC reporter”, who was actually an actor and had filmed the clip in 2016 in front of a greenscreen. The video spread wildly on WhatsApp, with many believing it to be true.²⁹</p>
LOOKALIKES/STAGED MEDIA	EXAMPLE
Lookalikes refers to the use of actors hired to mimic prominent figures. Staged media refers to instances of situations that are acted out, either by hired actors or average people, but presented as real. Both are difficult to detect without talking directly to the sources, and neither will be covered in this guide.	<p>In 2016, an actor was hired to mimic Senator Leila De Lima of the Philippines in a sex video, in order to undermine her credibility. The video was ultimately used as justification for her imprisonment.³⁰</p>
FACE SWAPPING, LIP SYNCHING AND VOICE SYNTHESIZING (DEEPAKES)	EXAMPLE
Face swapping, lip synching and voice synthesizing are examples of more advanced manipulation techniques. They can be considered deepfakes due to their high level of technological sophistication. For more information on the threats posed by this type of misinformation, see DRI’s deepfakes report. ²⁸	 <p>A deepfake video of world leaders singing “Imagine” by John Lennon.³¹</p>

28. Goldzweig and Brady, “Deepfakes: How Prepared Are We?”, op. cit., note 3.

29. Chris Baynes, “BBC Forced to Deny Outbreak of Nuclear War after Fake News Clip Goes Viral,” The Independent, April 20 2018, <https://www.independent.co.uk/news/media/bbc-forced-deny-fake-news-nuclear-war-viral-video-russia-nato-a8313896.html>

30. “De Lima on Sex Video: It Is Not Me,” Philstar Global, October 6 2016, <https://www.philstar.com/headlines/2016/10/06/1630927/de-lima-sex-video-it-not-me>

31. Mike Seymour, “Canny AI: Imagine World Leaders Singing,” Fxguide, April 12 2019, <https://www.fxguide.com/featured/canny-ai-imagine-world-leaders-singing/>

EXAMPLE
 <p>Deepfake video of former United States President Barack Obama with audio by Jordan Peele.³²</p>

Note: It can sometimes be difficult to tell whether manipulated media content is harmful or not. DRI supports the principle of free speech, including through the use of humour and satire, but there is a line that crosses over into harm. This line can sometimes be grey, so it is important for researchers to use their best judgment and understand the context of the misinformation to determine whether manipulated media content is harmful.

32. Craig Silverman, “How To Spot a DeepFake Like the Barack Obama-Jordan Peele Video,” BuzzFeed, April 17 2018, <https://www.buzzfeed.com/craigsilverman/obama-jordan-peeel-deepfake-video-debunk-buzzfeed>

3. Gathering Image and Video Post



How can visual content be collected from image- and video-based platforms?

When it comes to collecting manipulated media content, there are two potential methods — **targeted searching for specific events or messages, or the broad collection of thematic content**. Both methods have their pros and cons, and different platforms work better for different methods (see Table 2 for more detail).

In general, **targeted searching** is helpful when monitoring efforts have a clear goal of what they want to study and collect, or for those projects researching a specific event. For example, this could be a targeted search about a specific politician running in the Dutch Elections to find content posted in the months leading up to the elections about the candidate and their campaign.

Broad collection, on the other hand, is useful for research aimed at understanding general political themes. For example, as much data as possible could be collected related to the Dutch Elections, with a wider time frame and not about a specific politician, to find general themes and attitudes related to the event.

Table 3: Collection Methods

Collection Method	Pros	Cons
<div>Targeted Searches</div> <div></div>	<ul style="list-style-type: none">• Can filter content to collect only what's relevant• Less content to be organized and stored• Can be performed with no technical experience or programming skills• Uses only free, open-source tools	<ul style="list-style-type: none">• Time consuming• Requires significant human resources• There is the potential to miss relevant content• Less able to grasp wider narratives
<div>Broad Collection</div> <div></div>	<ul style="list-style-type: none">• Can collect a wide array of content with relative ease• Can be done by one person• Usually can be done relatively quickly• Able to grasp wider narratives	<ul style="list-style-type: none">• Often requires programming skills• Often uses paid tools• There is the potential to collect irrelevant content• More content to be organized and stored

This section will discuss YouTube, Instagram and TikTok – three popular image- and video-based platforms – and will provide guidance for both specific searches and broad collection.

YouTube

YouTube is the most popular video sharing platform on the internet.³³ As such, it is home to many instances of manipulated videos and is an excellent place to start collecting video content. YouTube also has a relatively user-friendly application programming interface (API), allowing researchers with some programming skills to collect large amounts of data with minimal work.

Specific Searches:

YouTube has its own search function, which can be used to search for videos using search terms. This is accessed by using the same search bar normally used to find a video. Researchers can use the built-in filters to search by upload date, type, duration and features. Results can then be further sorted by upload date, relevance, view count or rating.

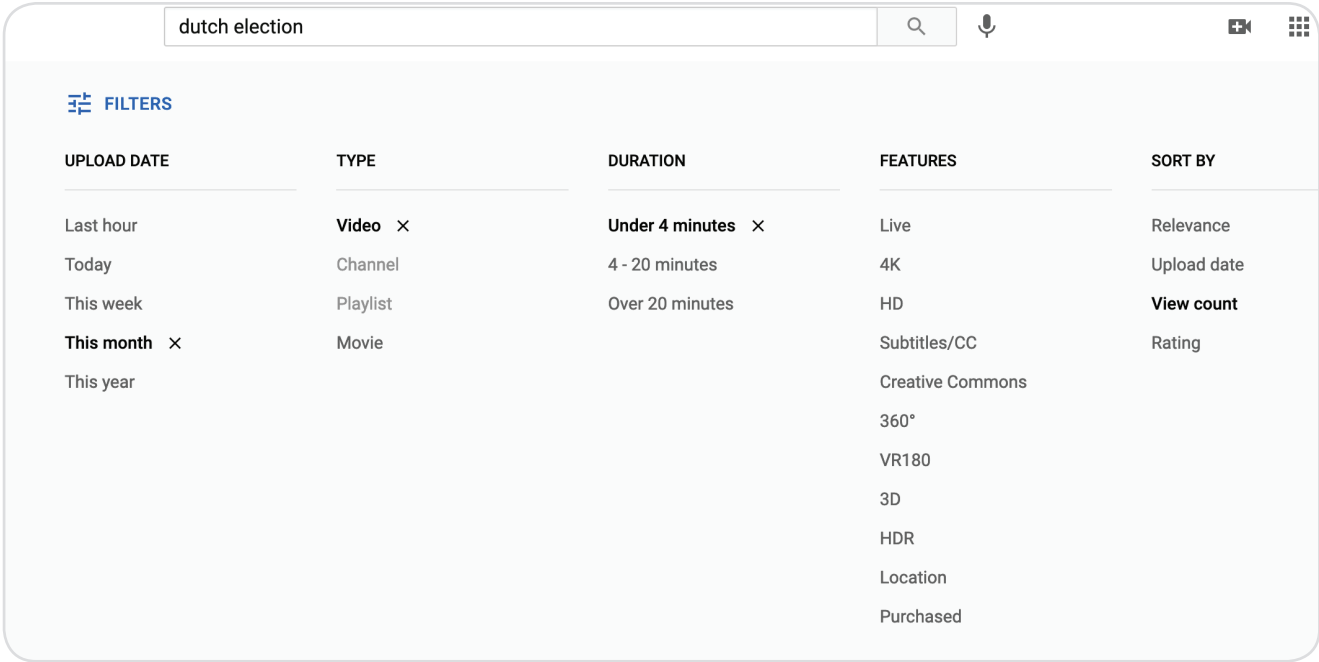
Figure 1: YouTube Search Filters

FILTERS				
UPLOAD DATE	TYPE	DURATION	FEATURES	SORT BY
Last hour	Video	Under 4 minutes	Live	Relevance
Today	Channel	4 - 20 minutes	4K	Upload date
This week	Playlist	Over 20 minutes	HD	View count
This month	Movie		Subtitles/CC	Rating
This year			Creative Commons	
			360°	
			VR180	
			3D	
			HDR	
			Location	
			Purchased	

For example, the appropriate search terms and these filters could be used to find short videos posted recently about the Dutch general elections, sorted by view count to find the most popular.

33. Haroon Malik and Zifeng Tian, “A Framework for Collecting YouTube Meta-Data,” *Procedia Computer Science*, The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017) / The 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2017) / Affiliated Workshops, 113 (January 1, 2017): pp. 194–201, <https://doi.org/10.1016/j.procs.2017.08.347>

Figure 2: Example Using YouTube Search Filters

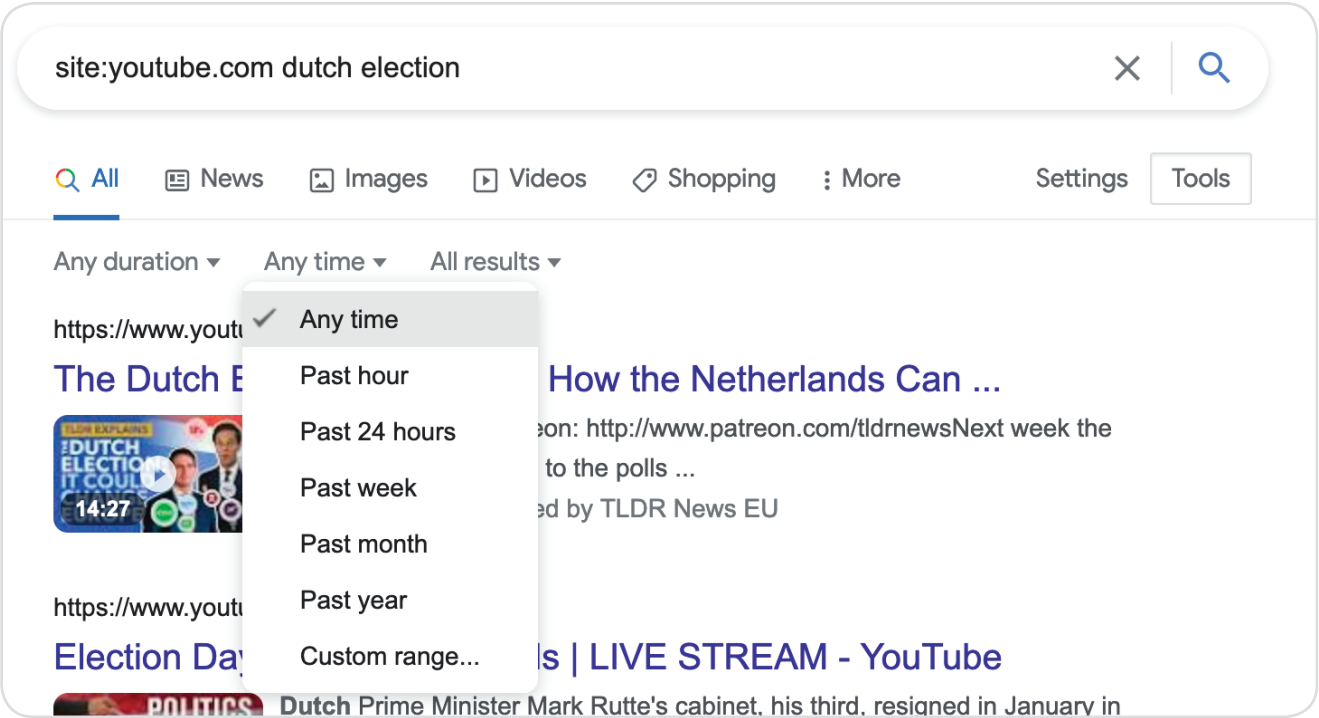


The YouTube filters are not very helpful however, for searching for a specific date or in a different language. An easy way to get around that is to use a site search on Google. This can be done simply by typing **site:youtube.com**, followed by the relevant search terms, into the Google search engine. From there, the results can be filtered by date and time by clicking “Tools”, and then “Custom range”.

Figure 3: YouTube Google Site Search

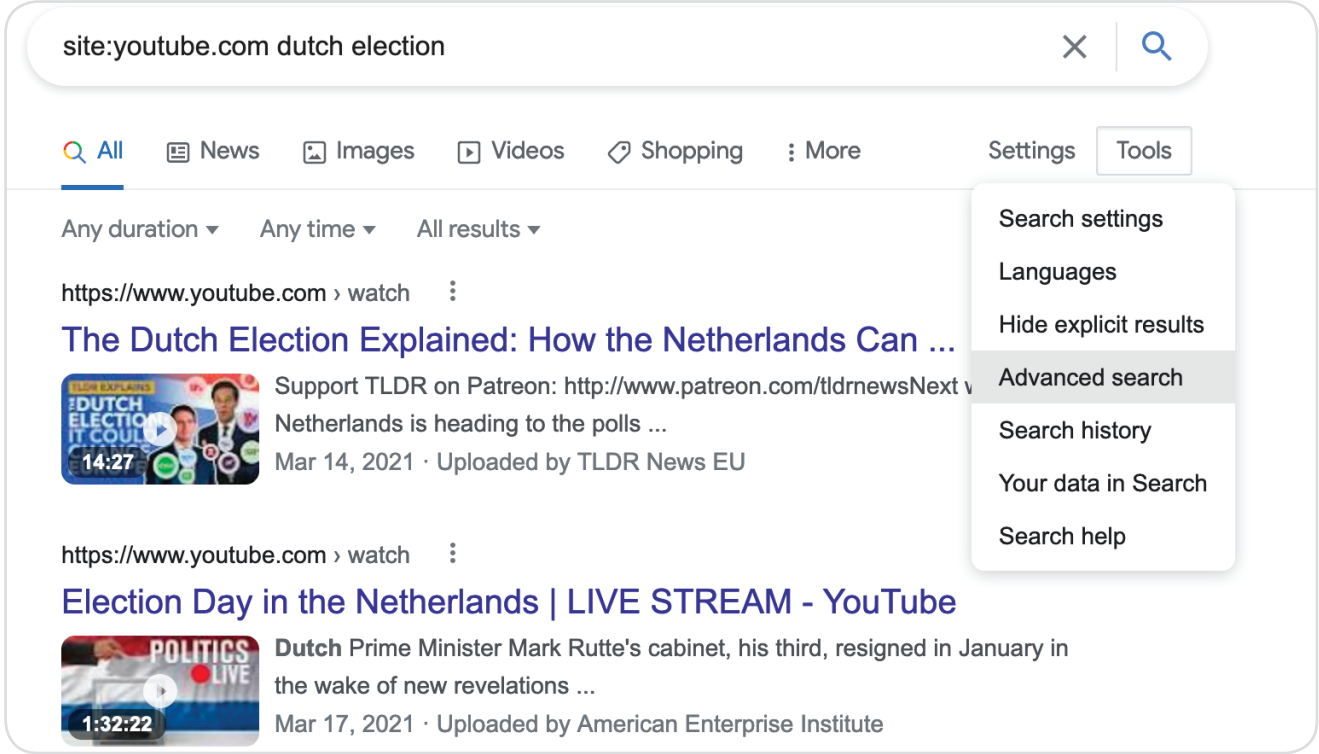


Figure 4: YouTube Google Site Search Results



Site searches can also be carried out using Google Advanced search, which allows for filtering by language, region, site or domain, and much more. This is done by simply clicking “Settings”, followed by “Advanced search”, and then specifying appropriately.

Figure 5: Google Advanced Search



Searches like these should identify videos relevant to the specific research topic or question. From there, the videos can be viewed to determine whether they are relevant, and then further analysed to see whether they contain misinformation. For more information on categorizing content, see page 42. For more information on analysing specific videos, see page 28.

Further Reading:

- [The Beginners Guide to Boolean Searches](#)
- [How to Conduct Comprehensive Video Collection](#)

Broad Collection:

As mentioned above, YouTube has an accessible API, meaning users can gain access to metadata for a wide swath of content for free. However, programming skills are required. YouTube data can be accessed using the steps detailed in DRI's guide to accessing data on YouTube.³⁴ The accessible metadata related to YouTube videos include views count, likes count, favourites count, video description and title, user comments and many other resources.³⁵ A tutorial for how to use YouTube API with Python can be found here, which includes Github code and other resources to help programmers get started.

Pulling metadata from YouTube can be helpful in a variety of ways. First, analysis of metadata can provide information about a specific YouTube account or video. For example, researchers at Peking University used the YouTube API to extract and analyse metadata about a specific user promoting conspiracy theories.³⁶ They were able to determine that the user had waged a sophisticated information campaign and created a web of all the accounts interacting with the original.³⁷ This technique can be useful for diving deep into accounts

API: An application programming interface is a set of functions and procedures allowing for the creation of software applications. In the case of social media research, APIs allow developers or researchers to access social media data. Their use requires programming knowledge.

Metadata: Metadata are, in general, data about other data. In the context of images and photos, this is information about the rights and administration of the image or video. This can include identification of the creator, of the model used to create the image or video, the date and time of creation and of posting, and a variety of other useful information.

34. "YouTube - Digital Democracy Monitor," Democracy Reporting International, <https://digitalmonitor.democracy-reporting.org/how-can-you-access-data/youtube/>

35. Ibid.; Haroon Malik and Zifeng Tian, "A Framework for Collecting YouTube Meta-Data", op. cit., note 33.

36. Ibid.

37. Muhammad Nihal Hussain et al., "Analyzing Disinformation and Crowd Manipulation Tactics on YouTube," in 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2018, 1092–95, <https://doi.org/10.1109/ASONAM.2018.8508766>

peddling misinformation related to elections. Second, it can be used to collect large amounts of data. For example, a research team at the ACM International Conference on Advances in Social Networks Analysis and Mining drafted a methodology that allowed their team to pull the metadata for over 700,000 videos, many more than could be accomplished through manual searching.³⁸

While these methods are helpful, due to their ability to collect large amounts of data with less effort, they require programming skills and time to set up the API scraping. If there are not the resources and/or time to go down this route, there are plenty of tools to help with manual searching listed in the section above, for YouTube, and below, for Instagram and TikTok.

Further Reading:

- [Analysing Disinformation and Crowd Manipulation Tactics on YouTube](#)
- [A Framework for Collecting YouTube Metadata](#)

38. Haroon Malik and Zifeng Tian, "A Framework for Collecting YouTube Meta-Data", op. cit., note 33.

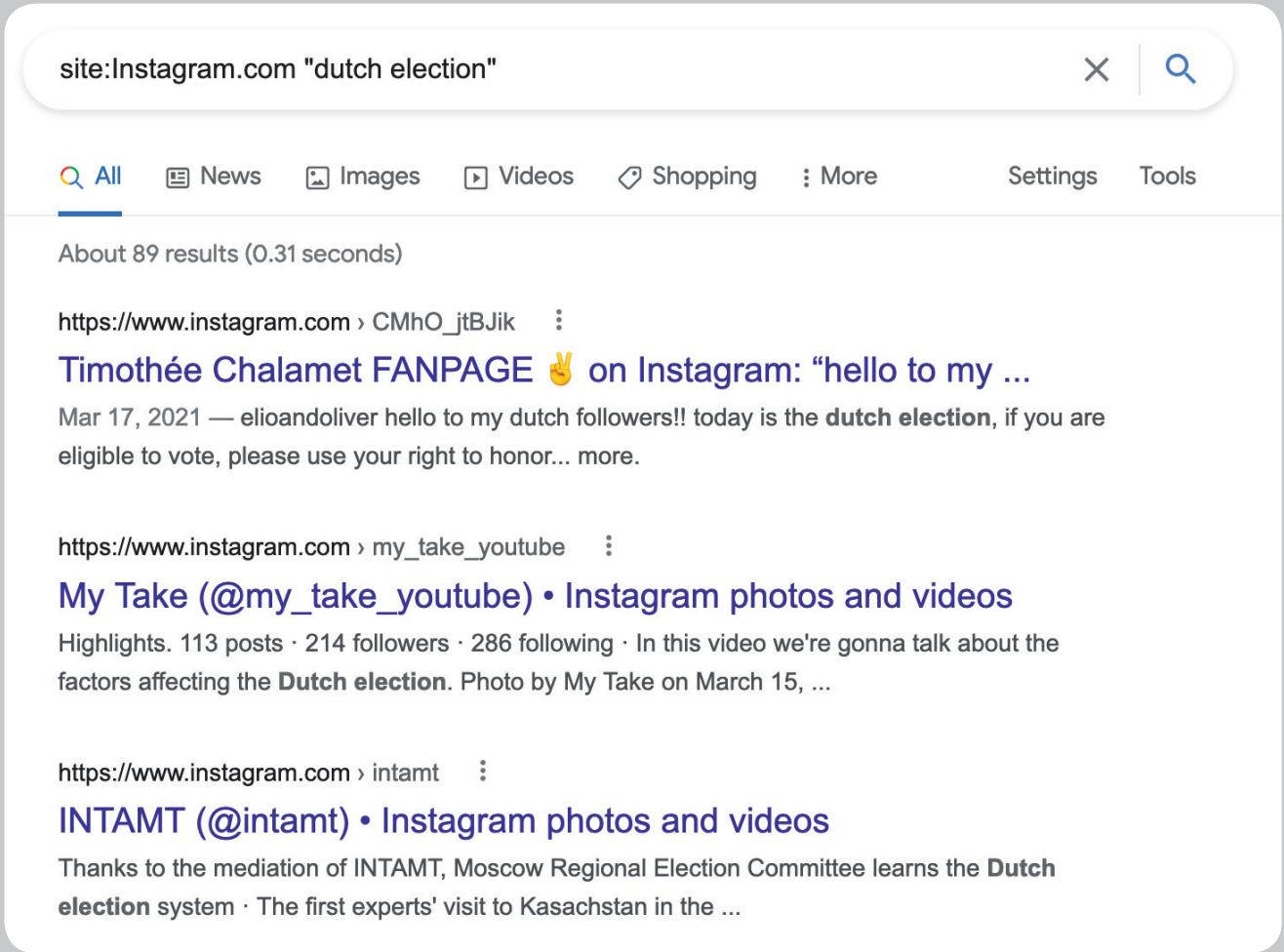
Instagram

Instagram is the world’s most popular image-sharing application. Owned by Facebook, Instagram is difficult to search. As such, it can be challenging to collect large amounts of data from this platform. However, there are some tricks and tools researchers can use to streamline the collection process.

Specific searches:

Similar to YouTube, Instagram has its own search function through its website, www.instagram.com, as well as through its mobile app. The search is quite clunky, however, and unlikely to produce helpful results. A Google site search, just like the one used for YouTube, is a better option for searching for specific content. A simple search, like the one shown below, will produce results of users with the search term in their bios, of those who have used the search term multiple times in their posts, or of specific posts containing the search term in the caption. These results can also be filtered by date using the Google search filters under “Tools” or by using the Google Advanced search features under “settings”. For more information on site searches, see page 14.

Figure 6: Instagram Google Site Search



Instagram also commonly uses hashtags. To search for a specific hashtag through Google site search, the hashtag must be put inside quotation marks. The first result displayed will show public posts made on Instagram containing that hashtag.

Figure 7: Instagram Google Site Search Results

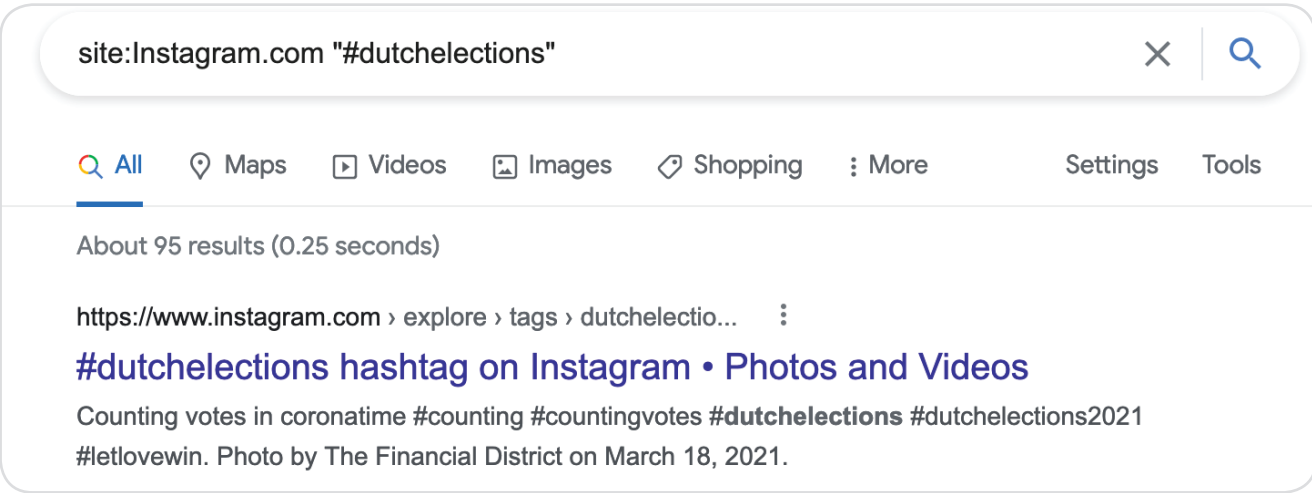
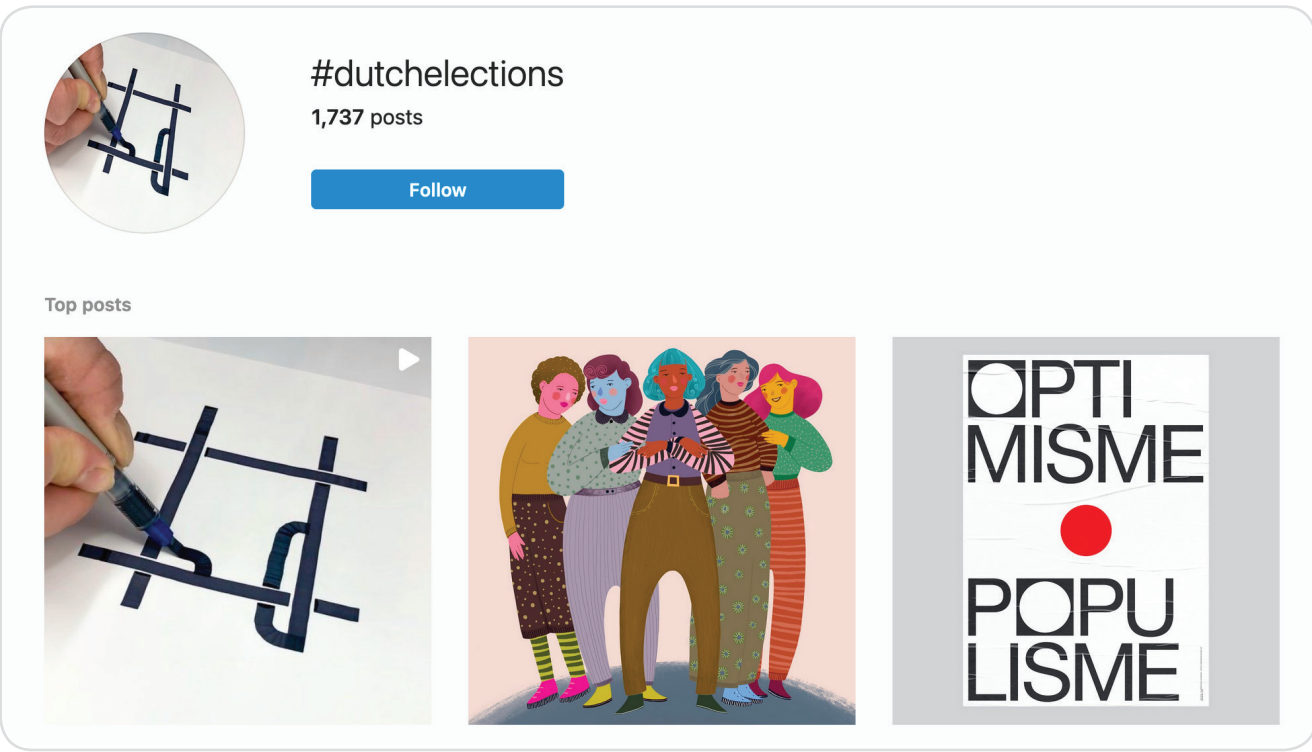
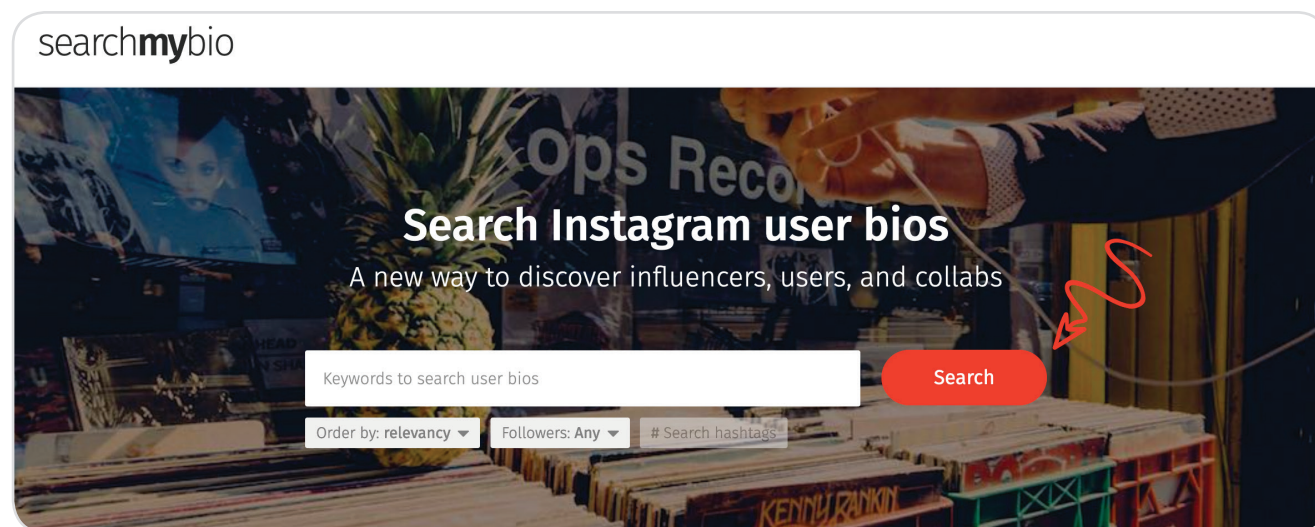


Figure 8: Instagram Hashtag Search Results



The Search My Bio tool can be used to search specifically for words in users’ bios. This site returns both private and public Instagram accounts with the chosen search term in their bio.

Figure 9: Search My Bio Example



Similar to YouTube searches, these techniques will result in a list of relevant image and video posts based on the search terms, and they will still require researchers to analyse and categorize each post. For more information on categorizing content, see page 42. For more information on analysing specific videos, see page 28.

Further Reading:

- [Searching Instagram by We Are OSINTCurious](#)
- [Searching Instagram Part 2 by We Are OSINTCurious](#)

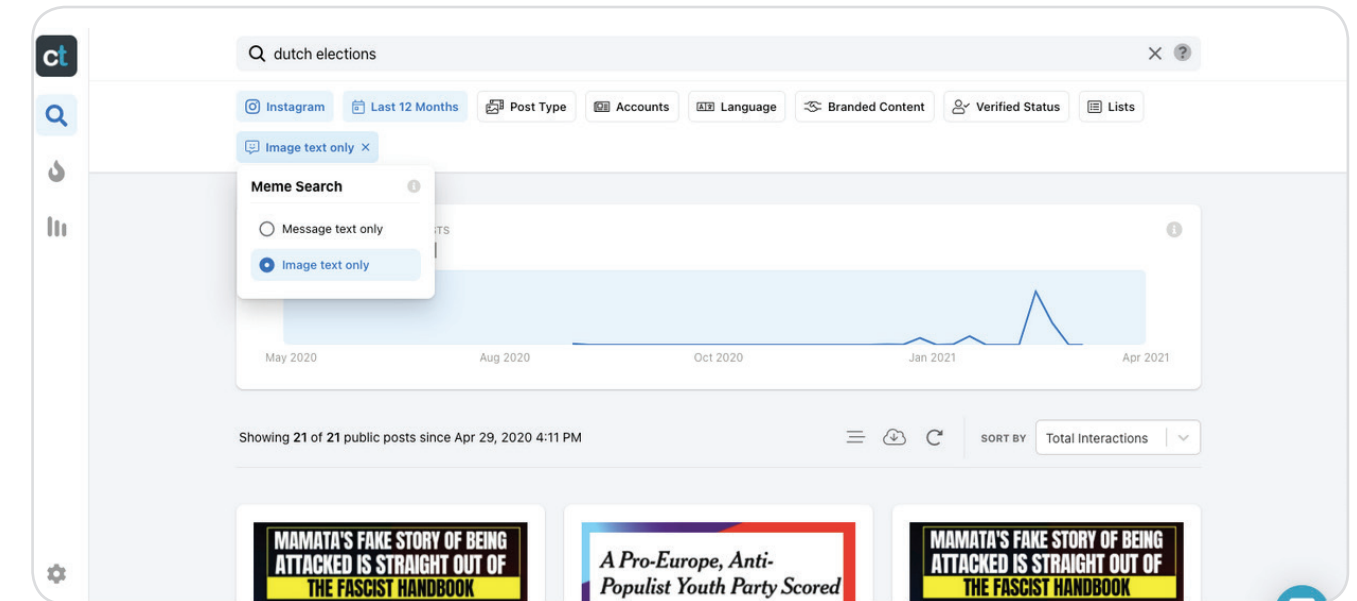
Broad collection:

As of April 2020, CrowdTangle was available for public Instagram data,³⁹ but only to Facebook partners, publishers, journalists, fact checkers and researchers. If a particular organization qualifies, access can be applied for [here](#).

For searching Instagram, the most helpful tool from CrowdTangle is their “Meme Search” feature. This tool allows users to filter through posts by both message text and image text. For example, when typing in “Dutch elections” as the keyword, CrowdTangle will return Instagram posts that mention this keyword within the image itself, as well as in the bio. Results can be filtered by date, post type, account language and various other metrics. It is then possible to download the results as a .csv file to analyse the filtered results further. The data downloaded includes total number of interactions, account name, number of followers and other helpful information.

39. Lauren Fraser, “What Data Is CrowdTangle Tracking?,” CrowdTangle, <http://help.crowdtangle.com/en/articles/1140930-what-data-is-crowdtangle-tracking>

Figure 10: CrowdTangle Meme Search



Further Reading:

- [Instagram: How can you access data?](#)

TikTok

TikTok is one of the fastest growing social media platforms today. It is also entirely image- and video-based. Since TikTok is relatively new, less research has been done about the platform and there are fewer tools available for collecting data. That being said, below are a few tips for accessing and collecting data from TikTok.

Specific searches:

TikTok is primarily a mobile application, although it has a desktop website as well. The information revealed from TikTok searches also depends on whether the user has a TikTok account.

Table 4: TikTok Mobile App versus Desktop Website

MOBILE APP:	DESKTOP WEBSITE:
Within the mobile app, there is a search function. This function allows for the use of search terms and filtering by top users, videos, sounds and hashtags. This, however, requires the use of either a smartphone or a virtual machine, both of which can be complicated to set up and come with a host of security issues.	TikTok.com also has a search function. It doesn't, however, always provide accurate or helpful results, and there is no filtering mechanism.

Besides the mobile application and website, there are a few other methods for searching TikTok. One is using Google site search, as with YouTube and Instagram. Other methods include searching for TikTok videos on other social media platforms. TikTok videos are commonly reposted on Facebook and Twitter, both of which have more user-friendly search functions.

Accessing TikTok Data on Facebook

TikTok videos can be searched for on Facebook by simply typing “tiktok.com” and the search term into the search bar, and then selecting “Videos” from the menu on the left side of the screen. From there, the results can be filtered by most recent or by date posted. Another option is to try “m.tiktok.com”, which will show videos shared directly from the TikTok app, or “vm.tiktok.com”, which will show videos shared using the “copy link” function from the TikTok app. Each will provide slightly different results.⁴⁰

Figure 11: Searching for TikTok videos on Facebook, example 1

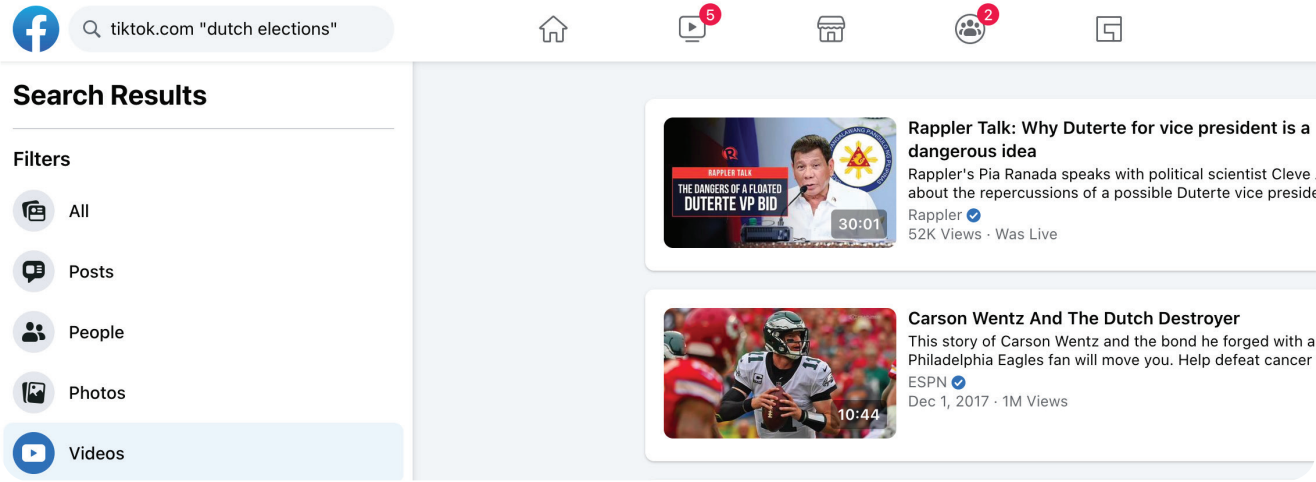
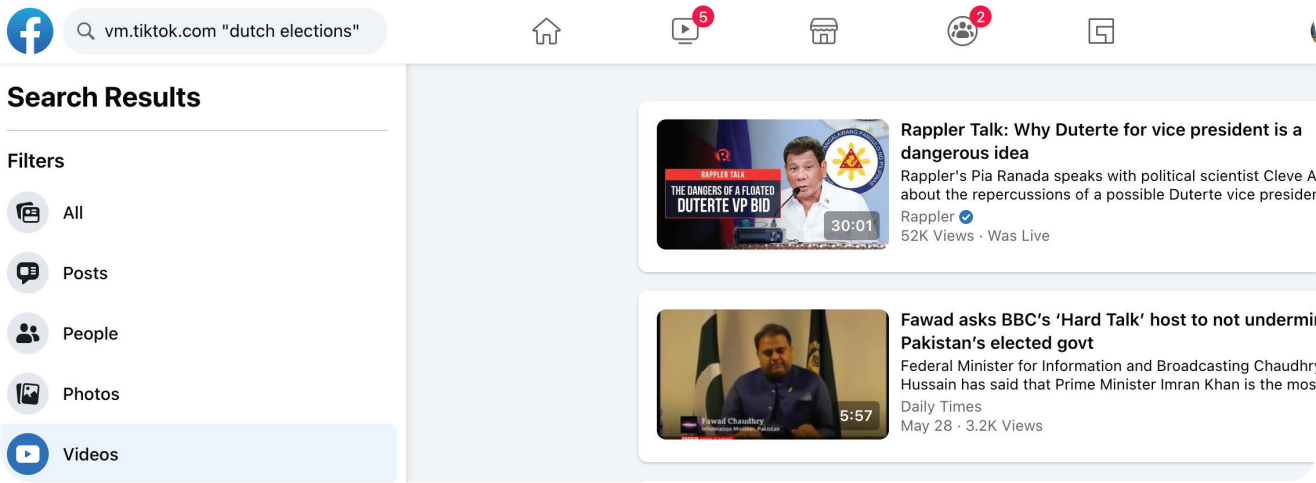


Figure 12: Searching for TikTok videos on Facebook, example 2



40. Annique Mossou and Johanna Wild, “Investigate TikTok Like a Pro!”, Bellingcat, May 25 2020, <https://www.bellingcat.com/resources/2020/05/25/investigate-tiktok-like-a-pro/>

Accessing TikTok Data on Twitter

TikTok videos are also commonly reposted on Twitter. The same “tiktok.com”, “vm.tiktok.com”, and “m.tiktok.com” searches can be used here, as well as the hashtag #TikTok, which is commonly found on videos reshared on Twitter.⁴¹

Broad collection:

Unlike more established social media platforms like Twitter and YouTube, TikTok does not currently provide official API access for researchers. There are various unofficial TikTok APIs, all of which require knowledge of programming. Accessing these scripts will allow a researcher to collect videos posted by a user or liked by a user, to generate a list of users, to collect trending videos, and more. Table 4 contains some resources and tutorials on how to collect data from TikTok, as well as links to GitHub packages.

Note: These resources have been created by online users and are not part of the official TikTok API. DRI has not tested these resources.

Table 5: TikTok Data Collection Methods

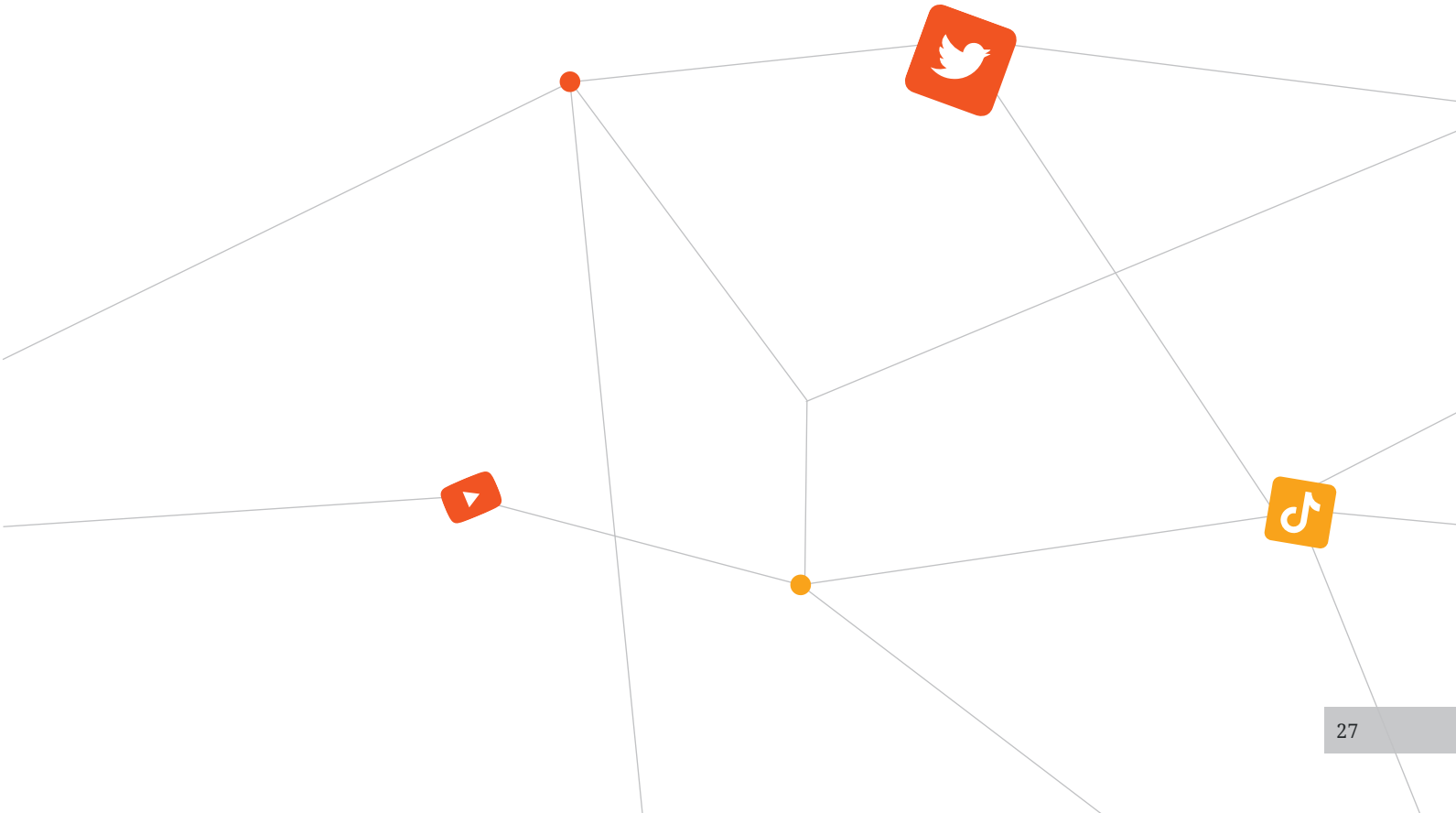
NAME	TYPE	PROGRAMMING LANGUAGE	FEATURES AVAILABLE	EXAMPLE
How to collect data from TikTok tutorial	Article on Medium.com, includes link to GitHub page	Python	Collect videos posted by a particular user; collect video likes by a particular user; use the information from a single profile to collect a list of similar users; collect trending videos	Can be used to easily collect all of the videos posted by an account known to be posting misinformation
TikTok API Python	Article on Medium.com, includes link to GitHub page	Python	Collect trending music; collect trending hashtags; collect trending TikTok videos	Can be used to collect all the hashtags related to a current election event
TikTok API	GitHub page	Python	Same as above	Same as above

41. Ibid.

TikTok R	GitHub page	R	Collect TikTok trends; collect videos posted by a particular user; collects TikTok videos by hashtag; collect TikTok videos from music ID; download TikTok videos	Can be used to easily download a TikTok of interest for further analysis and verification
--------------------------	-------------	---	---	---

Further Reading:

- [How to Create and Use Virtual Machines](#)
- [Investigate TikTok Like a Pro](#)



4. Analysis of Images and Videos

How can it be determined whether content has been manipulated?

How to Verify Individual Pieces of Content

When looking at a specific piece of visual content, whether it be an image or a video, there are several basic steps to determine whether the image or video includes misinformation or disinformation. The first few steps are useful for determining whether media content has been recontextualized, sped up, slowed or cropped, and are fairly straightforward and require less time and effort. The latter steps are geared towards investigating more advanced forms of manipulation, including photoshopping or otherwise editing images and videos. These steps and tools are more intensive, and should only be followed if the earlier steps point towards manipulation. This whole process is often called “verification”, as its goal is to verify that a piece of content is authentic.

Note: The examples used in this section do not necessarily contain instances of misinformation. Rather, they are used to demonstrate how specific verification tools can be used.

1. For a start: Trust your instincts:

If a piece of content feels like it is “off” somehow, this should be enough to trigger a closer look. Humans are better at detecting manipulated content than we think we are, and instincts are often an important first step in recognizing manipulated media content. This first step is an important one to come back to throughout the verification process.

2. Reverse Image Searching:

Reverse image searching is a powerful way to determine whether content has appeared in any other locations on the internet. This can help find the earliest instance where an image or video has been posted and is often the easiest way to find whether the image has been recontextualized. Reverse image searching can also sometimes be helpful in determining whether a video has been sped up or slowed down, as it can reveal original versions of the content. Occasionally, it can even point to research that has already determined whether certain content has been manipulated. Overall, it is an important early step for verification and analysis.

Reverse image searching allows the image to be used as a search term in various search engines. This can be done manually, by taking a screenshot and uploading that screenshot into the search engine of choice, but there are also a variety of tools that can be used to speed up the process. It is also important to remember to use a variety of search engines. Yandex, a Russia-based search engine, is known for having strong reverse image searching capabilities – stronger than those offered by Bing and Google. Overall, it is best to search the image in as many search engines as possible, and especially in the search engine that is most commonly used in the region of interest. Below are some tools that can be used to automate reverse image searches.

A. InVid

InVid is a verification tool that is very useful for reverse image searching YouTube video content. It is available as a [free plugin](#) for both Chrome and Firefox. Rather than manually screenshotting different frames in a video, InVid has a tool that allows a YouTube URL to be inputted, automatically returning reverse image searches on the chosen platform. For example, after inputting this 42 minute news report from Democracy Now, InVid automatically opens multiple Google reverse image searches from multiple points throughout the video.

Figure 13: InVid YouTube Thumbnails Reverse Image Search

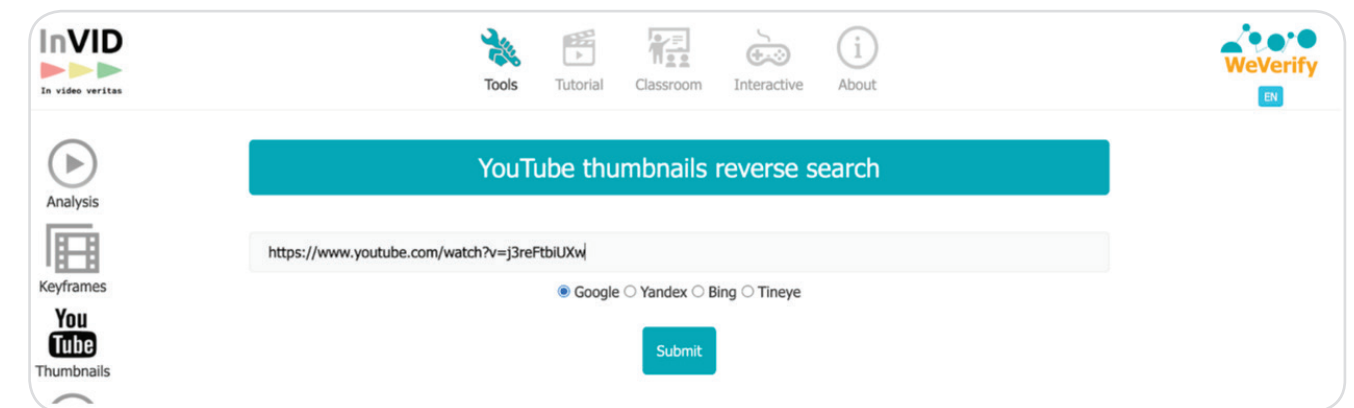
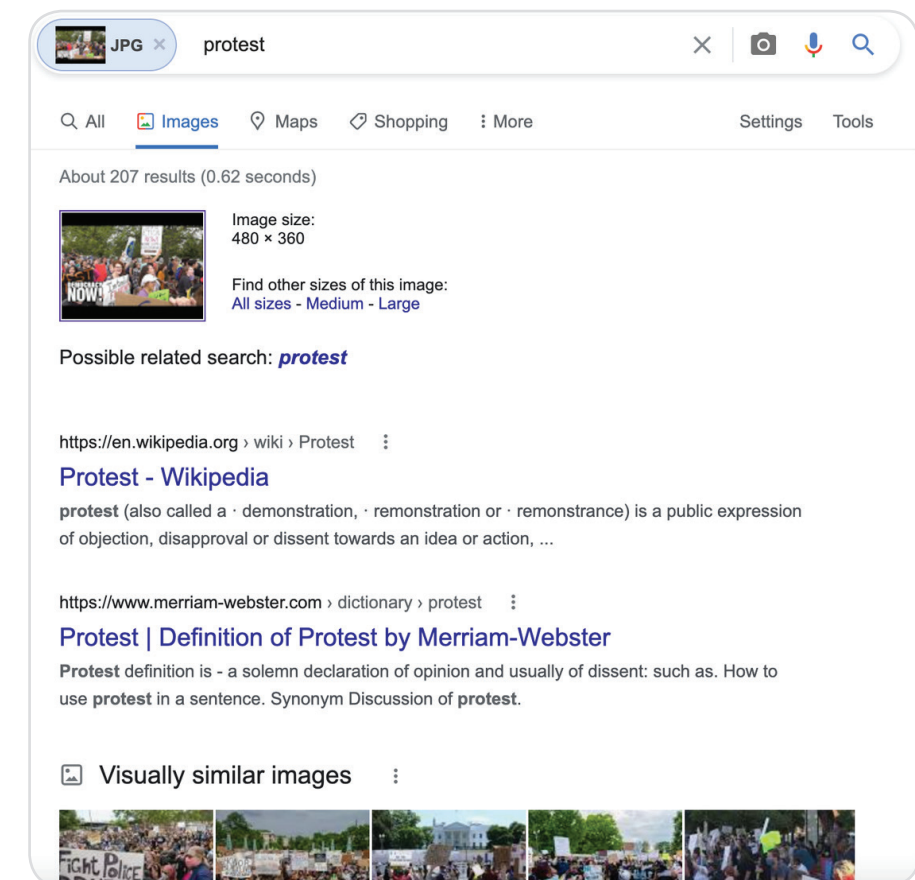


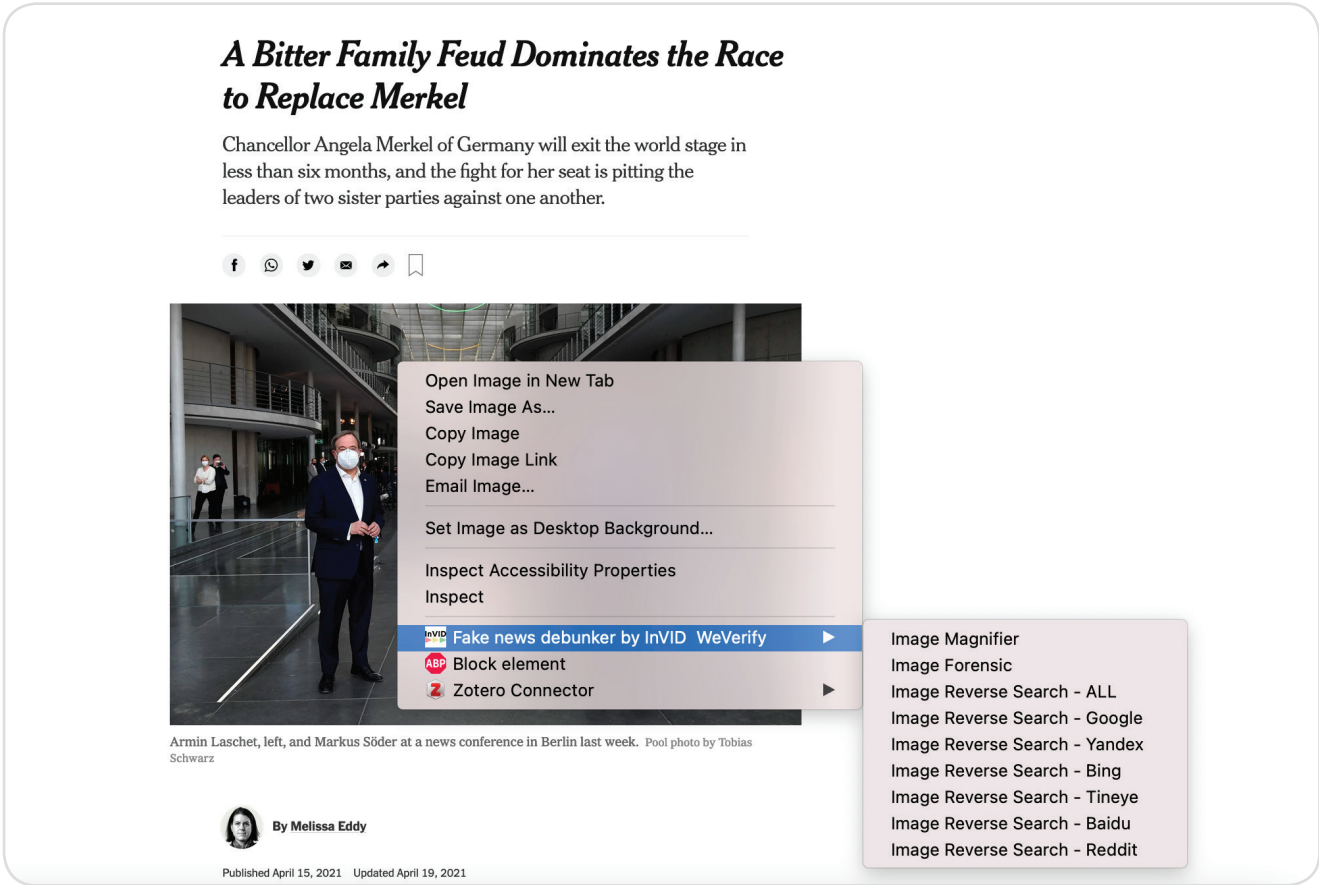
Figure 14: Google Reverse Image Search Results



InVid has another feature that allows for accessing more keyframes and choosing which to reverse image search. This can be accessed using the keyframes function. After inputting a YouTube link, InVid will return keyframes from throughout the video, which can then be easily reverse image searched.

InVid is also useful for reverse image searching images as they are found, by simply right clicking on the image and choosing which search engine to use.

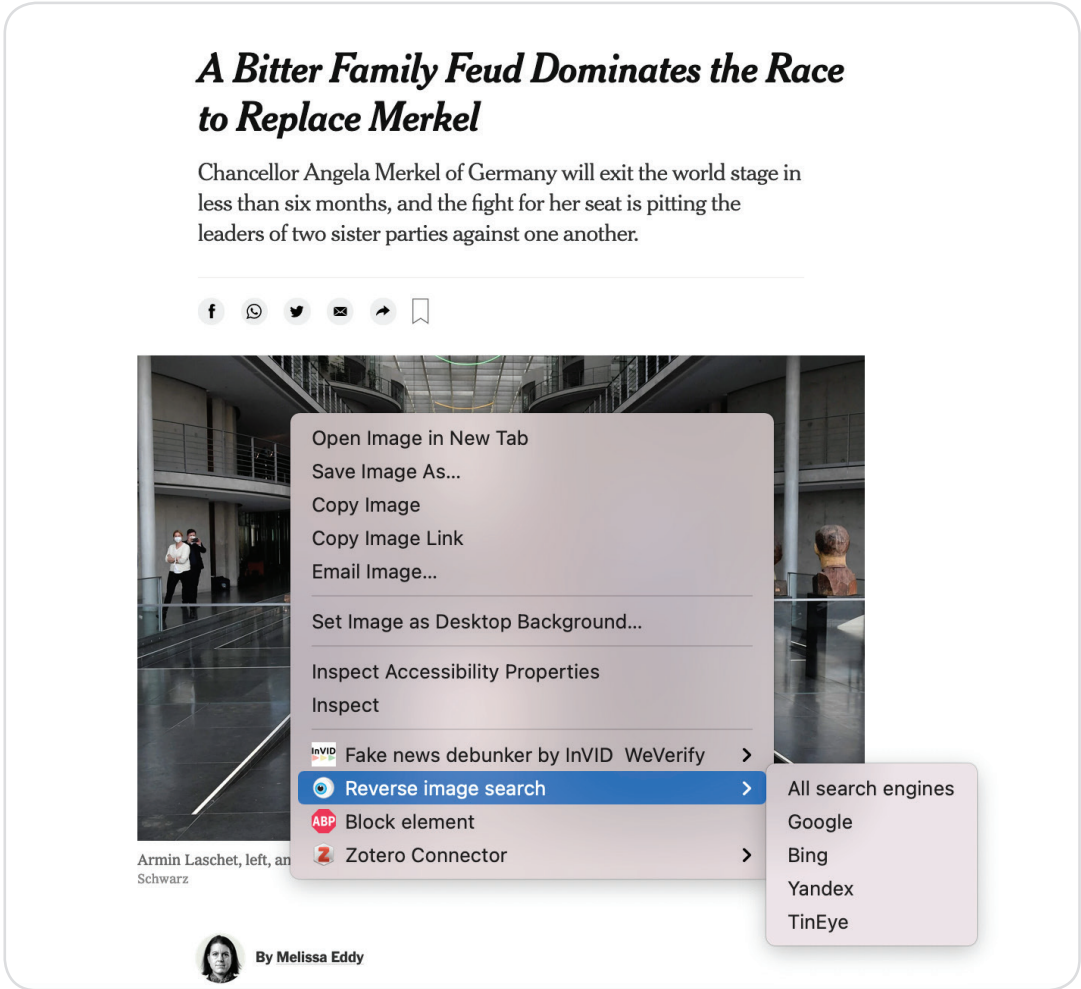
Figure 15: Invid Reverse Image Search Example



B. RevEye

RevEye is another useful tool for reverse image searching and is also available as a plugin for Chrome and Firefox. The steps are the same for reverse searching on InVid, by right clicking on the image and choosing a search engine.

Figure 16: RevEye Reverse Image Search Example



3. Investigate profile of user:

If reverse image searching doesn't automatically reveal that the media content has been recontextualized, the next step is to investigate the profile of the user who posted the content. Oftentimes, inauthentic accounts or bots are responsible for spreading manipulated media content. In other instances, users may be real, but investigating their account activity can reveal whether they commonly post manipulated media content, helping researchers to better understand the context of a specific piece of content. This investigation can be done manually, simply by visiting the profile of the user and scrolling through recent content and profile information. There are also a variety of tools available for investigating specific users on different platforms, allowing for a deeper dive into their activity.

Tip: a common piece of information to check is the location of the user. For example, if the piece of content shows alleged election interference in Ukraine but the profile of the user reveals that they are based in the United States, that is good indication that the piece of content may have been recontextualized.

A. Twitter

Twitter is a common platform for misinformation research and, as such, there are a variety of open-source tools available for investigating specific users.

Twitonomy: Twitonomy is an incredibly useful tool for analysing the activity of specific Twitter accounts. It reveals basic information, such as the number of tweets, followers and likes, as well as more advanced features, including the hours of the day when users post and what apps they post from. This information can reveal bot-like activity. For example, if an account is tweeting at all hours of the day, it is likely that it may be a bot. Twitonomy is free to use and requires only a Twitter account, but it sets limits on the number of searches. Twitonomy premium offers unlimited searches.

Botometer: Botometer is a tool that measures the likelihood of a Twitter user being a bot. Although it is not highly accurate, it provides a good starting point for investigating Twitter users. After entering a Twitter username, Botometer will return a score from 1-5, with 1 not being bot and 5 definitely a bot. It also scores the user on various categories, which can provide more specific information about a user’s activity.

B. Facebook

Facebook is another common platform for misinformation research. It is notoriously difficult, however, to search and find information, and tools used for profile analysis are often taken down. The best option is to use the site itself and search for specific users.

C. Instagram

Instagram is owned by Facebook and, as such, is also difficult to investigate. Below are a few tips and tricks for analysing profiles on Instagram.

Inspect element to view profile photos: This is actually something that can be done without any special tools. It works not just on Instagram, but on most websites and social media platforms. This trick can be used in order to view or download Instagram profile photos, which are often quite small. On Instagram, this is done by simply right clicking on the profile name, then clicking “Inspect” on a Mac or “Inspect Element” on Windows. This will open up code on the right side of the screen. From there, users should click control F and search for the phrase “img alt”. Next is looking for the link that follows, right clicking, and choosing “Open in new window”. This will open the profile picture in a new window.

Figure 17: Inspect Element Tool for Instagram Part 1

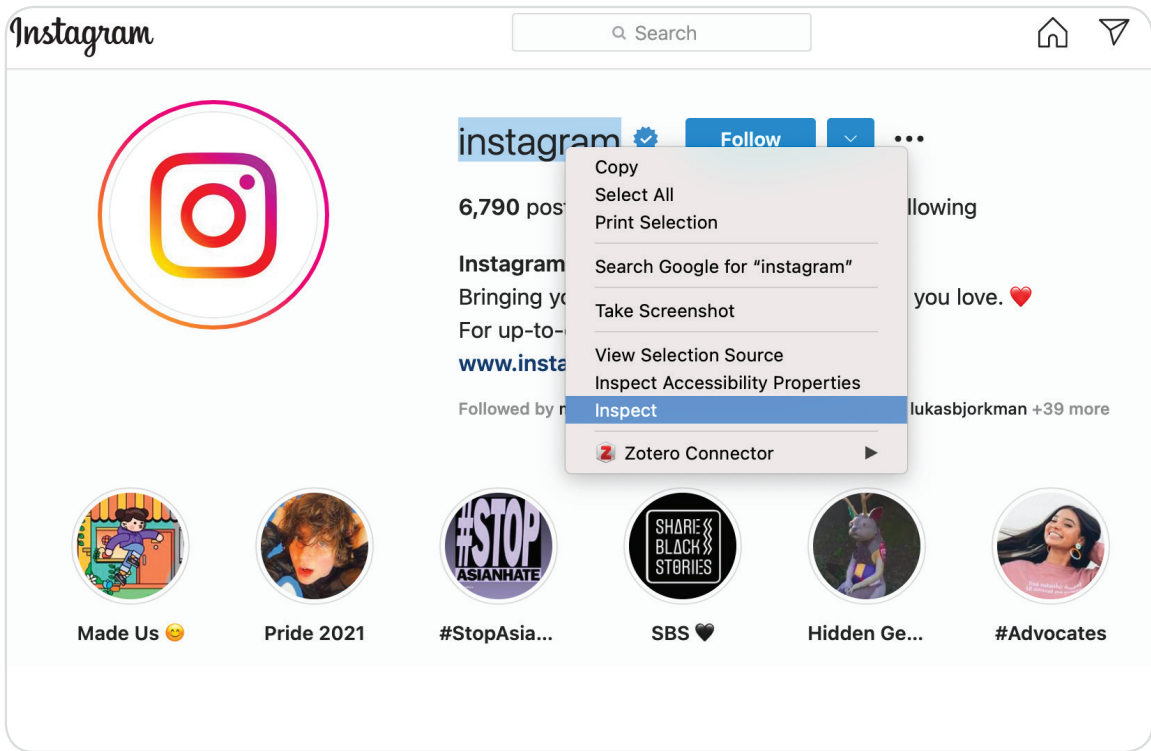


Figure 18: Inspect Element Tool for Instagram Part 2

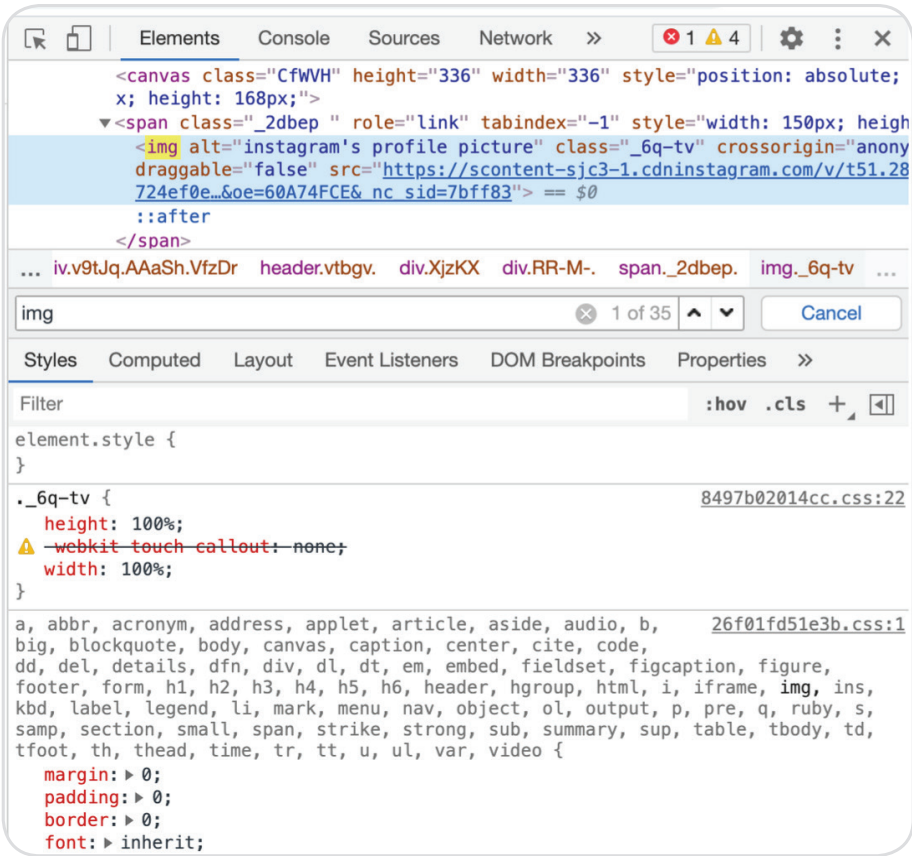


Figure 19: Inspect Element Tool for Instagram Part 3

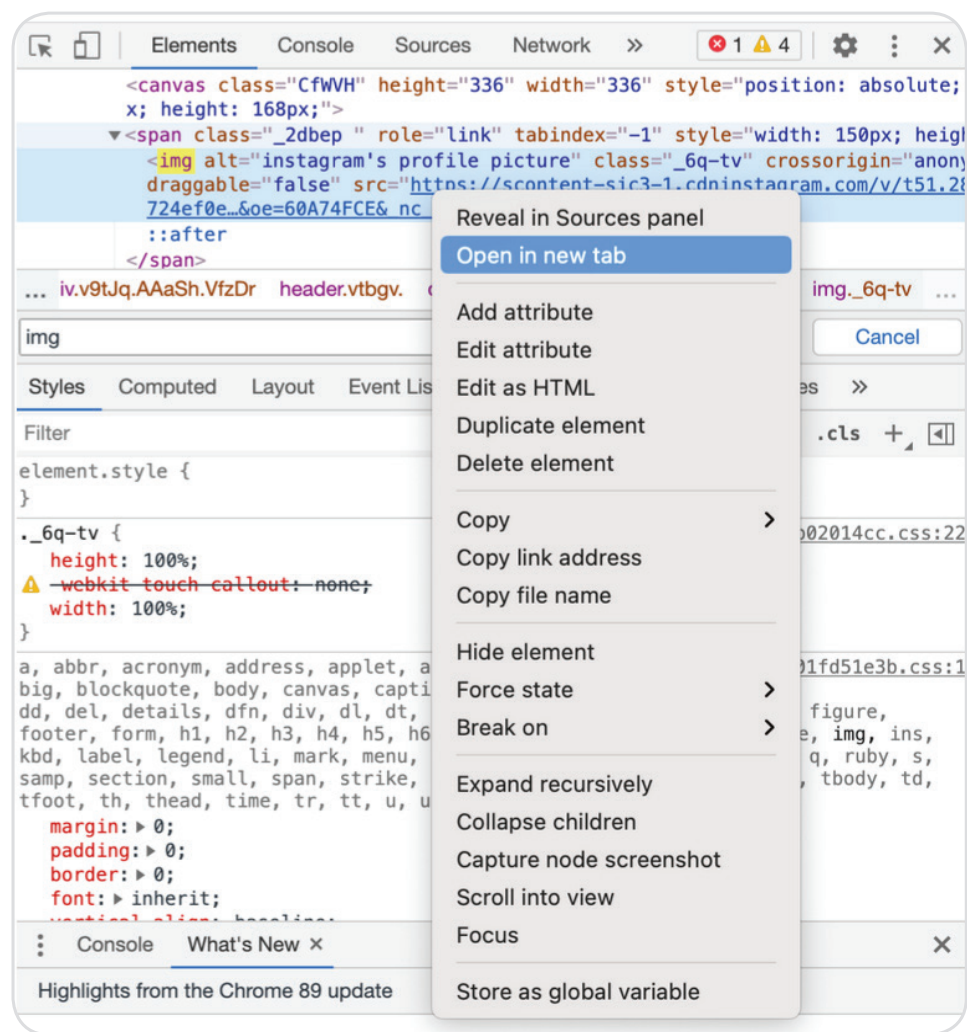


Figure 20: Inspect Element Tool for Instagram Part 4



D. YouTube

As the most common video-based platform, YouTube is an important one for verification and analysis. There are several tools available that can reveal information about specific channels, including the number of videos and subscribers, as well as about specific videos, including likes and comments.

[YouTube Data Viewer](#): This tool takes a YouTube URL as its input, and then immediately returns all of the metadata available for that video.

[InVid Metadata](#): InVid also has a function that takes a YouTube video URL, and then returns its metadata. This can be accessed through the same Chrome or Firefox plugin used for reverse image searching.

[YouTube Comments Downloader](#): This tool takes a YouTube video URL as its input and returns all of the comments and replies from that video, including the number of likes and dislikes. The site allows for searching within the comments, as well as for downloading the text data.

[Downsub](#): Downsub is a free online tool that allows users to download the subtitles from YouTube videos by simply inputting the URL. The languages available depend on the video, but the subtitles can be downloaded as either an SRT or an TXT file.

E. TikTok

As a relatively new platform, TikTok has fewer tools available for the analysis of specific users. Below are a few guides and tricks for investigating TikTok users.

[Bellingcat Guide](#) on investigating TikTok: This guide, published recently by Bellingcat, walks users through the basics of searching for content on TikTok, and has a section specifically geared towards investigating individual users.

Inspect element to view profile photos: The same “inspect element” trick discussed on page 32 for Instagram also works on TikTok, allowing researchers to better view profile photos.

Tip: AI-generated profile photos that look like real people are often used by inauthentic users. A common source is <https://thispersondoesnotexist.com/>. Some things to look for when trying to determine whether a profile photo is fake are strange or pixelated teeth, weird backgrounds, hair that looks unnatural, asymmetry or water splotches.⁴²

4. Geolocation and chronolocation:

Investigating manipulated media content sometimes requires determining where and when the event in the image or video took place. If, for example, a video claims to show election interference by police, it is important to confirm whether or not the events in the video actually took place at the date and time it claims. Geolocation is the process of determining where an image or video was taken, and chronolocation

42. Jevin West and Carl Bergstrom, “Which Face Is Real?,” <https://www.whichfaceisreal.com/learn.html>

is the process of determining when it was taken. Full training on geolocation and chronolocation is beyond the scope of this guide, but Table 5 contains a few tips and tricks, as well as links to other resources.

A. Look at comments, captions, reposts and likes.

Users will sometimes comment on the location of an event, providing a place to start. Information about the user can also help narrow down where the image or video may have been taken. A quick Google search about the event can also sometimes be helpful in providing information about the time and location of the event.

B. Reverse image search

It can be helpful to do another reverse image search when geolocating, in order to find other instances of the piece of content that might contain information related to time and location. It can also sometimes reveal similar images, which could provide more helpful angles of the same location.

C. Looking at buildings, signs, roads, license plates, stores, etc.

These items will be incredibly helpful in providing a place to start. Sometimes, a video will even show a street sign, providing a clue as to where it was taken. Shop names and road signs are also helpful.

D. Utilize satellite imagery

To fully verify the location of an image or video, satellite view should be used to match items in the content to items viewable from satellite imagery.

Table 6: Geolocation Resources

Geolocation Resource	Description
Free Amnesty International Training Course	A thorough guide to open-source investigations, with chapters specifically dedicated to geolocation and chronolocation.
Chronolocation Guide from Bellingcat	A guide that walks readers through the basics of how to determine when an image or video was taken, by using the sun and shadows.
Watch Frame by Frame	An online tool that allows users to easily watch YouTube videos one frame at a time. Helpful for both geolocation and chronolocation.

5. Deep dive into manipulation detection tools:

If the previous steps don't reveal any obvious manipulation, but instincts still suggest that something has been altered, the next step would be to look more deeply into the image or video, this time searching for more advanced alterations (tools for detecting deepfakes will be discussed later on). There are several helpful tools available for use in deep image and video verification and analysis.

Fotoforensics: Fotoforensics is a free online tool that allows for Error Level Analysis, or ELA. ELA shows

areas within an image that are at different compression levels, and can be used to detect manipulation of an image. It is important to note that ELA is not 100 per cent accurate; depending on how many times the image has been saved or resaved, the ELA will become less and less accurate. Because of this, it is important to try to find the first iteration of the image, using reverse image search, in order to minimize the number of times it has been saved. ELA has also not been scientifically researched as a forensic method but, rather, has been developed as a tool for amateurs or hobbyists.

Table 7: How to use Fotoforensics

Reverse image search the image of interest to find the earliest possible version.
Upload the earliest image into Fotoforensics.com (either an image link or a file).
Click on the ELA tab on the menu on the left of the screen.
Use the ELA evaluation tutorial provided by Fotoforensics to analyse the ELA image provided. Remember, ELA is not exact but, rather, an accessible tool to check for obvious manipulation.

[Image Verification Assistant](#): A free, online tool from Reveal that analyses images using five different methods. Images can be uploaded from a URL or a local file. Each method is explained on the website, and various examples of manipulated media content are provided.

[Fake Video News Debunker by InVid](#): InVid also has a function allowing for image forensics. It offers the same five verification methods as the Image Verification Assistant from Reveal.

Further reading:

- [Verification Strategies in Visual Journalism](#)
- [‘Fake’ World Press Photo isn’t fake, is lesson in need for forensic restraint](#)
- [YouTube Data Tools](#)
- [What Can Newsrooms do to Tackle Manipulated Media?](#)
- [Guide to Using Reverse Image Search for Investigations by Aric Toler](#)
- [Automatically Reverse Image Search Videos by Justin Seitz](#)
- [Data Craft: The Manipulation of Social Media Metadata](#)

How to Detect Deepfakes

Deepfakes have drawn increased media attention in recent years. In interviews with DRI, several human rights experts told us that they are generally able to verify image and video content using the techniques described in the sections above. As artificial intelligence and deep learning advance, however, the possibilities for the creation of increasingly realistic deepfakes also grows. Luckily, this advancing technology also provides more resources for those working to detect deepfakes, allowing them to create tools for combating misinformation and performing political analysis. Below are some existing tools on the market to detect deepfakes, with information on their availability and their use. They should only be used if there is any reason to believe that AI was used in the creation of the content in question. See DRI’s report on deepfakes for more information. (Note: Details on the capabilities of the tools based on information from the respective developers).⁴³

Sensity: Sensity has four functions: deepfake detection, search, threat assessment and watchlist. With deepfake detection, users can input a file or URL and Sensity will analyse the content to detect manipulation. The search function allows users to access a comprehensive database of deepfakes. The threat assessment provides information on specific public figures’ susceptibility to visual threats. Finally, with the watchlist, users can set up notifications for when new visual threats are detected.

Sentinel: With Sentinel, users can upload content through their website. Sentinel will analyse the image or video to determine whether it is a deepfake.

FakeNetAI: FakeNetAI has both a web app and an API that can scan videos for manipulation. They provide free trial access.

Quantum + Integrity: According to their website, Quantum + Integrity detects all levels of manipulated media content and can be tailored to specific needs. Beyond deepfake detection in video, they can also perform identity, insurance and document verification.

Microsoft Video Authenticator: Although not yet available to the public, the tool will be able to determine, by detecting elements that are not visible to the human eye, the chance that an image or video has been manipulated, expressed as a percentage.

Although addressing the threat of deepfakes can sometimes be daunting, the work of researchers to combat these advanced manipulations provides hope for the future of political monitoring.

Further Reading:

- [Machine Learning Researchers Spot Deep Fakes from Heartbeats](#)
- [Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches](#)
- [Protecting World Leaders Against Deep Fakes](#)
- [FaceForensics++: Learning to detect Manipulated Facial Images](#)
- [Recurrent Convolutional Strategies for Face Manipulation Detection in Videos](#)
- [Deepfakes and beyond: A Survey of face manipulation and fake detection](#)
- [Deepfakes: How prepared are we?](#)
- [Deep Learning for Deepfakes Creation and detection: A Survey](#)

43. Goldzweig and Brady, “Deepfakes: How Prepared Are We?”, op. cit., note 3.

How to Monitor on a Large Scale

In the context of an election, media content is often being posted at astounding rates. For an organization attempting to monitor elections, it can be overwhelming to sift through all of this content. There are two main solutions to the problems of large-scale monitoring. The first is to employ more people. An example of a successful large-scale monitoring effort is the Human Rights Center at UC Berkeley’s monitoring project for the 2020 United States presidential elections, which involved over 60 people.⁴⁴ Involving more people allows for more content to be collected, categorized and archived.

Of course, involving more people is not always an option. There are also various automated options available. Many involve programming skills to pull information from various social media platforms using their APIs (for more resources on this, see earlier sections on the YouTube, Instagram and TikTok APIs). There are, however, also tools available to help with large-scale media monitoring that require no programming skills.

TweetDeck: TweetDeck allows for multiple advanced Twitter searches in columns across a researcher’s screen. The columns update automatically, which is an excellent feature for active monitoring. For more information, see this [TweetDeck guide](#) from Bellingcat.⁴⁵

Note: The following tools are paid web scrapers. They all automate the web scraping process and require no programming experience. Some have free trials available. DRI has not tested the effectiveness of these tools.

Dexi.io: An automated web data extraction software that can be personalized to the needs of the user. Free trial available.

Octoparse: An interactive tool that allows users to quickly extract data from a web page, with no programming required. Free trial available.

Outwit Hub: Options for the user to create their own web scraper or to have Outwit Hub create one for them. Online sales had been suspended as of April 2021.

Parsehub: A free desktop app that allows users to download data from specific websites. Data access is via JSON, Excel or API. No programming required.

Scraping Robot: Built for developers, this tool allows users to scrape websites into JSON. Free trial available.

Zyte: Formerly Scraping Hub, Zyte allows for data extraction at scale. Free trial available.

44. Maeve Sneddon and Andrea Lampros, “Live Monitoring Election 2020,” Human Rights Center, November 7 2020, <https://medium.com/humanrightscenter/live-monitoring-election-2020-da061439d05f>

45. Charlotte Godart, “The Most Comprehensive TweetDeck Research Guide In Existence (Probably),” Bellingcat, June 21 2019, <https://www.bellingcat.com/resources/how-tos/2019/06/21/the-most-comprehensive-tweetdeck-research-guide-in-existence-probably/>

What else is possible with deep learning?

In addition to creating tools for detecting deepfakes, the advancement of AI technology has rapidly increased the availability of off-the-shelf deep learning tools. These are largely being shared in online code-sharing platforms such as GitHub, where users can adapt prewritten code to their needs. Several examples are shared below.

The YOLOv3 tool, which uses a convolutional neural network for object detection, including in real-time, can be run with a few lines of code once downloaded, with image tagging done via Microsoft's user-friendly Visual object Tagging Tool (VoTT). This technology allows researchers to automatically detect objects of interest in images, for example, to find a gun in a large collection of images.⁴⁶

In 2020, Analytics Vidhya created a meme detector tool using machine learning technologies. This tool can take an image and detect whether or not the image is a meme – a challenging feat given the complexity of memes and their usage. Although the tool is not yet available for use, its implications for the field of deep learning for political analysis are impressive.⁴⁹

Optical character recognition, or OCR, is an artificial intelligence technique that recognizes text from images. OCR has become ubiquitous and easy to access, showing how such technology can become widely applicable in a short amount of time.⁴⁷

Vision AI by Google is an API with a variety of applications. It can detect faces, emotions, landmarks, crops, image properties, labels, landmarks and logos, as well as web properties, including where else the image exists on the web. There is currently a version that requires no programming skills but provides limited capabilities without payment, as well as a more advanced version that requires some programming. This tool has the potential to make verification of manipulated media content much easier and more streamlined, and it is likely only a matter of time before there are many more similar, even more user-friendly tools available.⁴⁸

Researchers from Binghamton University and Intel Corporations released a paper in 2020 introducing the use of machine learning to detect deepfakes using heartbeats. According to their research, their method is able to detect deepfakes with over 90 per cent accuracy.⁵⁰

In addition to studying the fingerprints of GANs, GANs have also recently been developed to detect other GANs. For example, in 2019, researchers from the University of Washington and the Allen Institute for Artificial Intelligence developed a GAN called GROVER, which generates fake media content in order to better detect it. This can be applied to both textual and visual news, with some limitations.⁵²

Generative adversarial networks (GANs) are deep learning models that allow computers to analyse a set of data and generate a new example based on patterns they encounter.

An example of this is <https://thispersondoesnotexist.com/>, which generates false images of human faces (see page 35). Recent studies have shown that GANs leave online fingerprints, meaning that researchers can track who created them, thereby monitoring coordinated misinformation campaigns by nefarious actors.⁵¹

Further Reading:

- [Surveillance using facial recognition and social media data by Eriksson et al.](#)
- [Facebook is using billions of Instagram images to train artificial intelligence algorithms](#)
- [5 Compelling Social Media Scraping Tools Available on the Web](#)
- [An Overview of Textual and Visual Content to Detect Fake News](#)
- [Fake News Detection: A Hybrid CNN-RNN Based Deep Learning Approach](#)
- [Detection of GAN-Generated Fake Images Over Social Networks](#)

46. Anton Muehleemann, "How to Train Your Own YOLOv3 Detector from Scratch," October 4 2019, <https://blog.insightdatascience.com/how-to-train-your-own-yolov3-detector-from-scratch-224d10e55de2>.

47. Karrar Haider, "7 Ways To Convert Images to Text Using OCR," Geekflare, February 14 2021, <https://geekflare.com/convert-image-to-text/>

48. <https://cloud.google.com/vision>, <https://cloud.google.com/vision/docs/detecting-web>

49. Chouaib Nemri, "Create and Deploy Your Meme Detector from Scratch : A Step-by-Step Guide," Analytics Vidhya, January 10 2020, <https://medium.com/analytics-vidhya/create-and-deploy-your-meme-detector-from-scratch-a-step-by-step-guide-3f5a96f22099>

50. Umur Aybars Ciftci, Ilke Demir and Lijun Yin, "How Do the Hearts of Deep Fakes Beat? Deep Fake Source Detection via Interpreting Residuals with Biological Signals," ArXiv:2008.11363 [Cs], August 25 2020, <http://arxiv.org/abs/2008.11363>; Ram Sagar, "Machine Learning Researchers Spot Deep Fakes From Heartbeats," Analytics India Magazine, September 10 2020, <https://analyticsindiamag.com/deepfake-heartbeat-machine-learning-detection/>

51. Ning Yu, Larry Davis and Mario Fritz, "Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 7555-7565, doi: 10.1109/ICCV.2019.00765.

52. Vincy Davis, "GROVER: A GAN That Fights Neural Fake News, as Long as It Creates Said News," Packt Hub, June 11 2019, <https://hub.packtpub.com/grover-a-gan-that-fights-neural-fake-news-as-long-as-it-creates-said-news/>

5. Storing Image and Video Files

How can manipulated content be stored and organized?

Organizing Content

One of the most difficult parts of dealing with visual content is categorization. When looking into visual content, it is not usually as simple as placing it into the categories of “true” or “false”. The categories will largely depend on the context being researched, and categories can and should change as the research progresses and as new information becomes available. as an example, the Facebook fact-checking categories are a good place to start.⁵³

Facebook fact-checking strategies:

1. False

- This is content that has no basis in fact. This includes, but is not limited to, impossible claims and fake quotes, or to image, audio or video content that is presenting itself as a separate event.
- Example: During the Black Lives Matter protests in the summer of 2020, a photo was circulated showing a McDonald’s on fire, claiming it was caused by the protests. In reality, the photo came from a McDonald’s that had burned down in Pennsylvania in 2016.⁵⁴

2. Altered

- This refers to content that has been doctored, edited or manipulated in ways that could deceive the viewer.
- Example: In May 2019, a video of United States Speaker of the House Nancy Pelosi was slowed down, making it appear that she was slurring her speech. The slowed video was popularized by multiple right-wing news sources.⁵⁵

3. Partly false

- This content contains some factual inaccuracies.
- Example: In 2017, then-United States President Donald Trump tweeted a claim that 1.5 million people attended his inauguration, when most other estimates were much lower.⁵⁶

4. Missing context/misleading

- This content may mislead viewers without additional information.
- Example: In early 2020, a video circulated on TikTok that showed then-United States Presidential candidate Joe Biden stating that “We can only re-elect Donald Trump”, leaving out the full quote, which was “We can only re-elect Donald Trump if, in fact, we get engaged in this circular firing squad here.”⁵⁷

5. Satire

- This refers to content that uses absurdity or irony, but that may be misconstrued as truthful by the average viewer.
- Example: A satirical article by The Babylon Bee, a well-known satirical website, was misconstrued as true and fact-checked.⁵⁸

6. True

- Content that contains no inaccurate or misleading information

Some other helpful categories in an election context could include “unverified”, referring to content that has not been determined to be true but has also not been debunked, as well as a category for memes, which often require different analysis than images and videos. Categories will need to be determined based on the research goals of each team.

Coding content can also be time consuming. With image and video content, more manual labour is required to categorize than with text data. With this in mind, there are a couple tips and tools that can make categorizing content easier.

53. “Rating Options for Fact-Checkers,” Facebook Business Help Center, accessed May 10, 2021, <https://www.facebook.com/business/help/341102040382165>.

54. Janet Lytvynenko and Craig Silverman, “We’re Keeping a Running List of Hoaxes and Misleading Posts about the Nationwide Police Brutality Protests,” BuzzFeed News, June 5 2020, <https://www.buzzfeednews.com/article/janetylytvynenko/hoax-misleading-claims-george-floyd-protests>; Chris Asroff, “McDonald’s East Complete Loss after Grease Fire,” Lebanon Daily News, November 18 2016, <https://www.ldnews.com/story/news/local/2016/11/18/mcdonalds-east-complete-loss-after-grease-fire/94103656/>

55. “Video: Edited Pelosi Video vs. the Original: A Side-by-Side Comparison,” The New York Times, May 24 2019, <https://www.nytimes.com/video/us/politics/100000006525055/pelosi-video-doctored.html>

56. Timothy B. Lee, “Trump Claims 1.5 Million People Came to His Inauguration. Here’s What the Evidence Shows.,” Vox, January 21, 2017,

<https://www.vox.com/policy-and-politics/2017/1/21/14347298/trump-inauguration-crowd-size>

57. Alex Kaplan, “TikTok Is Hosting Videos Spreading a Deceptively Edited Biden Clip, despite the Platform’s New Anti-Misinformation Policy,” Media Matters for America, March 10 2020, <https://www.mediamatters.org/fake-news/tiktok-hosting-videos-spreading-deceptively-edited-biden-clip-despite-platforms-new-anti>

58. R. Kelly Garrett, Robert Bond and Shannon Poulsen, “Too Many People Think Satirical News Is Real,” The Conversation, August 16 2019, <http://theconversation.com/too-many-people-think-satirical-news-is-real-121666>

Tips:

Categories should be created early on, without being afraid to have too many (it is easier later to combine categories than to split them up). Everyone who will be categorizing has to understand each category and its definition.

Categorization should be carried out throughout the process. This requires a clear methodology to be set up beforehand (i.e., spreadsheet with checkboxes for each category)

If the work will be carried out with data scientists or if the data are intended to be useful to them or the larger scientific community later on, it is worth consulting them when creating categories. The manner of categorization and classification can influence the types of analysis that can later be carried out easily.

Tools:

[AirTable](#): AirTable is a sophisticated tool for organizing and analysing data. Users are able to customize every part of the viewing and visualization experience to best fit the needs for their data. Something like Excel, but more customizable and with more features.

[Maltego](#): Maltego is an open-source intelligence tool that allows users to map connections between parties of interest. It offers both real-time data collection and mining and representation of this information as social webs. It is very helpful in making connections and finding patterns within large datasets.

Archiving Content

Storage of image and video files is an incredibly important part of any investigation. Posts containing misinformation are often removed or censored by social media platforms, meaning they may disappear in the course of research. To mitigate this, it is best to store any images or videos that seem relevant to the research project. Below are a few tips and tools for storage.

Tips:

Only things that fit into the categories created should be stored. Time or bandwidth shouldn't be wasted downloading content that won't be useful. This can often be a fine line, so it requires some judgement.

If possible, a few members of the team should focus only on archiving. This allows for a cleaner, more efficient workflow.

Tools:

[Hunchly](#): Hunchly is a tool designed for online investigations. It functions as a Google Chrome extension that tracks all of the activity in the course of investigations and saves any and all posts or content discovered. Users can tag pages and posts as they come across them, essentially organizing content as they go, and Hunchly will automatically archive everything. A free 30-day trial is available.

[Internet Archive Wayback Machine](#): This open source website allows any user to input URLs, which are then stored on the internet. It is an excellent place to archive images and videos without taking up

computer space. Users need to be aware that anything they input into the Wayback Machine will also be available to other users.

[Go Full Page Chrome Extension](#): Sometimes the best option is to simply screenshot the content in question. This extension allows users to take a screenshot of an entire web page with only a single click. It captures all content, even if the page requires scrolling.

[4k video downloader](#): This online tool allows for the easy downloading of images and videos from YouTube, TikTok and other video platforms.

[Instagram Video Downloader](#): To use this online tool, simply input an Instagram URL. It will return the downloaded video.

[Twitter Video Downloader](#): This Chrome extension allows Twitter videos identified to be downloaded with just one click.

6. Exercises

Below are a few exercises designed to help researchers practice the tools and techniques detailed in this guide. Most of the questions in these do not have just one correct answer, as results will vary depending on search terms used and the location of the search.

1. Try searching for YouTube videos related to the 2021 Benin elections. How many can you find posted on April 11th, 2021 – the day of the elections? What search terms can you use? What happens when you switch the search language to French, the official language of Benin?
2. Try searching for Instagram posts containing the hashtag #stopthesteal, a hashtag related to the 2020 United States presidential election. What kind of content do you see? Are there any prominent users using these hashtags? Are there any hashtags that are often used in conjunction with #stopthesteal?
3. Try searching for TikTok videos related to the protests in Chile in 2019, using both the Facebook and Instagram methods. Use the same search terms and dates and see whether the results differ. Which site provides more content? What video is the top result on each site? If you do the same search on TikTok itself, what happens?
4. Reverse image search the following image. What is the earliest iteration of this image you can find? Who posted this photo erroneously?



5. Choose your favourite YouTube video and upload it into InVid. What Metadata can you extract? What does it tell you about the video? About the channel?
6. Use [Botometer](#) to analyse the popular Twitter account @EmojiMashupBot. Does it correctly define this account as a Bot? How about its followers? Does Botometer define your Twitter account as a Bot?
7. Find your favourite celebrity's Instagram account. Use the “Inspect element” tool to open their image in another tab.
8. Using any tools you think might be useful, can you find the exact location where the following photo was taken?



(Answer: 51.50489030842836, -0.12635377597894543)

9. Scroll through your own Facebook or Twitter feed. Try to put every post you see into the Facebook categories listed on page 42. What complexities do you notice? Are there any categories you feel are missing? Are there any you don't need?

