

# AI breakthrough: A guide to tackling disinformation



# About Us!



[www.cri.lt](http://www.cri.lt)



CIVIC RESILIENCE INITIATIVE



@CivicResilience



@CRI



@civicresilienceinitiative

The publication is produced in cooperation with the Baltic Security Foundation (Latvia) and the National Centre for Defence and Security Awareness (Estonia).

This publication is sponsored by Google.

# AI breakthrough:

## A guide to tackling disinformation

Over the last ten years, the diffusion of digital media continues to proliferate. Every day, we are exposed to an increasing amount of information streams coming from a wide variety of communication channels: social networks, blogs, websites, traditional media, or other electronic publications.

This diversity of information streams makes it easy to choose the source that best reflects our interests and political or social views. As engagement and time spent on social networks grows more and more people are choosing them as their main source of information, often without appreciating the threats that exist within them. The rapid and convenient dissemination of information on social networks creates the perfect conditions for the rapid spread of disinformation.

The current popularity of artificial intelligence (AI) and the tools that rely on it provide ample opportunities for spreading disinformation and malicious information. The Internet of Things (IoT) - enabled video and audio fakes can cause significant damage to society by spreading fake news and thus undermining citizens' trust in the media. However, with great potential

for malicious provocation comes the equal potential to counter it. Once IoT has been mastered, the ingenuity of this technology can also be used in the fight against disinformation.

**Lithuania is a constant target of false information, and malign actors who actively seek to undermine its credibility.**

Such efforts fuel distrust of local government, our partners in NATO and the European Union, and continuously try to demotivate citizens to actively participate in governance. While professionals do a great job of debunking false stories, often the damage is already done once the disinformation has been disseminated.

To educate the public in the never-ending fight against disinformation, in 2019 we founded the Civic Resilience Initiative (CRI). The organisation collaborated with disinformation and media experts and came up with the idea to develop a practical guide,

"AI Breakthrough: A Guide to Tackling Disinformation". **This disinformation identification toolkit offers a simple guide to the methods needed to verify information.**

We hope that this publication will help you easily develop the habit of checking the authenticity of the news sources you read and the photos and videos that accompany them.

**The CRI team seeks to be the main catalyst in Lithuania and the Baltic region to strengthen the digital resilience of our societies.**



**This practical guide to tackling disinformation will give you the basic tools to help you:**

- check whether information on the internet is real or fake;
- understand how the IoT works and how to recognise images that have been manipulated by AI;
- identify trolls and fake social media accounts;
- recognise fake images and videos on the internet;
- take action when you see disinformation;
- warn others about falsehoods spread on social media.

This toolkit aims to contribute to increasing digital resilience and raising awareness about security and the need for vigilance in the information space.

**The aim of this publication is to provide information to help students, and the general public to strengthen their digital literacy skills and their resilience to hoaxes and false information.**

The publication is produced in cooperation with the Baltic Security Foundation (Latvia) and the National Centre for Defence and Security Awareness (Estonia).

*This publication is sponsored by Google.*

# There are **many different** forms of fake news:

## Disinformation

Information that is false and deliberately designed to harm an individual, social group, organisation or country.

## Misinformation

Information that is false, but not created with the intention of causing harm.

## Malinformation

Information that is based on reality, used to inflict harm on a person, organisation or country.

All three types of information are dangerous because they travel far and fast and because they go 'viral': this happens when many people, and even organisations, repost these stories because they seem interesting and sensational, without giving them much thought.



# Identification:

## How to verify news or posts?

Reading a scandalous news story with a topic or content that sounds unbelievable? Step back and check the credibility of the information. What should you do?

### Here are five simple steps, to verify the information:

---

#### 1. Rate the source

Explore the website or social media account. Think about who might be behind the distribution of the news and what was the purpose of the story.

---

#### 2. Read past the headline

The headlines of stories can be scandalous, and are used to attract clicks and promote sharing. If you dwell into the story, it may turn out that the claims in the headline are not true.

---

#### 3. Check out the author

Is the author a reliable person? Does the named author even exist?

---

#### 4. Do the sources confirm the story?

Often, there are no links in fake news that can be used to verify the facts. If there are references to sources in the story, then click through them. It may become apparent that the original message has been embellished or the meaning distorted.

---

#### 5. Check the date

Re-publishing old news does not mean that they are still relevant.



## Additional tips:

### Choose safe and reliable information.

For news in Lithuanian, use major portals such as LRT, DELFI, 15min, Alfa, TV3 and similar. Smaller portals may be easier to influence to buy certain content due to lower human and financial resources, have fewer journalists to check the credibility of the information, and may be more vulnerable to cyber-attacks and hacking.

#### Carefully filter information originating on foreign portals or social media groups:

- Your sources must have a sufficient following. Otherwise, look for who else has published the same information and filter those people again.
- If you collect information on X (formerly known as Twitter), TikTok, or YouTube platforms, comments must be enabled on trusted channels. More trustworthy entries will have more than one comment, with solid gaps between postings, and will not be bound by writing style templates. Otherwise, comments may be made by trolls or bots.
- Sources should not have links to pro-Russian government portals (Sputnik, PBK, Rossiya 24, and others). Although there may be some truth in the news published by these portals, in the context of an information war, it is not worth taking the risk of believing Russian news sources unless they are confirmed by the portals of Lithuania or its foreign partners.
- Information that is sensational should always be based on reliable sources. If you see breaking news on the internet, make sure that the source of the breaking news also mentions its own sources. If not, check with Google:
  1. Add quotation marks to the most important keywords from the post or article you're reading and type them into Google search.
  2. Click on the "Tools" button on the right edge of the search box and select to show the latest.

3. Check the sources again against the above criteria.

- Ask yourself if this source is really lacking in evidence and analysis. From which perspective is the news presented? Could it be the case that the events being reported are false and that there is an attempt to manipulate the imagination and emotions of the readers?

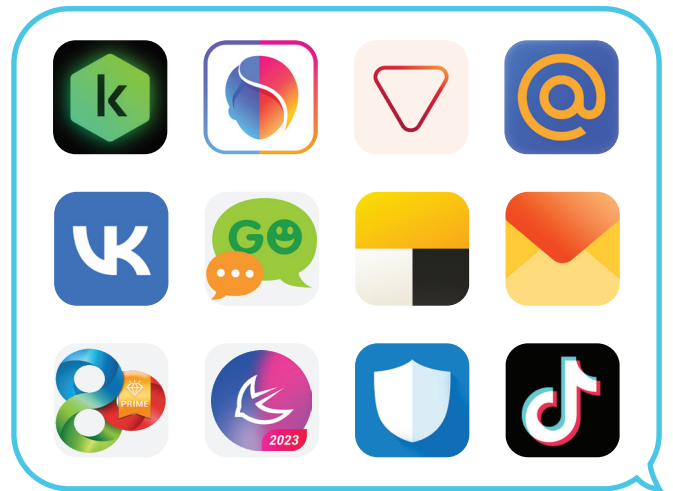
Finally, when you see a sensational news item on one of the websites of one of our public institutions (e.g. the Ministry of Foreign Affairs, [urm.lt](http://urm.lt)), visit a reputable news portal. One of the tools of information warfare is the hacking of government websites and the dissemination of misleading information on them. If our public institutions are hacked, the media will certainly report it. Then we should not believe the information they publish. In some cases, hacking is not necessary and fake website addresses may be used. For example, [urm.com](http://urm.com) would replace [urm.lt](http://urm.lt) or, more cleverly, [urm.it](http://urm.it) (i instead of l). Such tricks are hard to spot, so always check that the address line really looks like it should.

### Offer trusted sources of information not only to your parents but also to your grandparents.

Do your family members watch Russian TV? Military information has been one of the most popular and most discussed topics on Russian TV channels in the last few years. Understandably, for many of our parents or grandparents, Russian is the main and preferred foreign language. We would therefore suggest that your family members install Netflix or another streaming platform that includes Russian translated content. As Russian television is designed to make the viewer apathetic to information, the transition to a home cinema platform should not be difficult.

## Uninstall Russian and Chinese apps.

Uninstall any apps produced in Russia or those from their potential allies, such as China, from your phones. Some of them have various accesses to your data or location, which can be used against you either to spread disinformation or in military action. The most popular apps include: Kaspersky Antivirus, CheckScan, FaceApp, MyPocket, Mail.ru, V Kontakte, Go SMS Pro, Yandex Taxi, Yandex Mail, Go Launcher, APUS Launcher, Security Master and even TikTok.



# Generative Artificial Intelligence

(AI)

Generative AI is associated with artificial intelligence models that can generate content such as images, text, music, etc. Examples of generative AI include well-known models like ChatGPT, DALL-E, and others, where we can easily input text queries and receive generated text or images in return.

Large language models (LLMs) are a type of generative AI that can understand, recognize, contextualize, and generate text. To enable LLMs to perform their tasks, models need to be trained, and this training requires a substantial amount of textual data. Typically, data is obtained by reading open source text from the internet and transforming it for training purposes.

Language models are often used to answer questions that we do not know the answer, generate written text when we do not have time to prepare, or even write programming code when we are not proficient

in programming languages. This is an excellent tool for obtaining information or creating a short piece of content that would otherwise require a significant amount of time and effort to produce.

However, LLMs are not only used for positive purposes but also enable cybercriminals to conduct cyber-attacks much more effectively:

### 1. Social engineering

LLMs can create excellent text in any language, and this text will be more accurate than translations done through the „Google Translate“ program. The text will be of high quality with few or no errors. Therefore, a cybercriminal may not know the language of their chosen victim but can still write text that attempts to trick the victim into giving money or clicking on a link that tries to coerce them into providing login information. As a result of LLMs, malicious emails,



messages, and texts will be of higher quality and harder to identify as fraudulent. There are dedicated language models specifically designed for generating such deceptive messages.

## 2. Spreading misinformation

The response generated by LLM programs is not always accurate and may not reflect reality, leading users to be misled, and receiving an inaccurate or wrong answer can begin to spread within the user's social circles. Additionally, models learn from user questions and the clarifications provided by users, so the model may learn "facts" that are incorrect and present them as accurate to others, contributing to the spread of misinformation.

# Generative Artificial Intelligence (AI)

## How to recognise text generated by large language models:

Currently, there is no tool that can definitively determine whether a text is written by artificial intelligence or not. The most accurate recognition tool currently is

the human brain. For example, the text below is written with the assistance of ChatGPT. **Will you spot the errors in it?**

Dear User!

I trust this email finds you in good health. I am writing to bring your attention to the importance of completing the registration/login process on our platform.

As part of our commitment to providing you with a seamless experience, we kindly request you to finalize your registration or log in to your account at your earliest convenience. This will not only ensure the security of your account but also grant you access to the full range of features and benefits available on our platform.

To complete the registration/login process, please follow the link provided below: Login Link

If you encounter any difficulties or have any questions, please do not hesitate to contact our support team.

Thank you for your cooperation, and we look forward to serving you.

This short text has some areas, that looks suspicious:

- The first sentence does not sound natural, and nowadays it is rare to receive an email with such a beginning. Usually, such start suggests that text is generated.

- Also, the first paragraph is written from a first-person perspective ("I"), however the remaining text is written from a "we" perspective, which also looks suspicious.

Hence, this example illustrates that artificial intelligence, when generating text, uses phrases which are more common to automatically generated text, and the inconsistent pronoun usage raises questions. Other clues may be related to the use of incorrect cases or sentence structure errors. Additionally, sometimes the language's culture and style may reveal the work of artificial intelligence, even if there are no errors in the text.

There are several websites online that attempt to determine whether a text is generated by artificial intelligence or not. To understand how automated tools try

to identify whether the text was written by AI, we need to go back to the basics of how LLMs generate text.

In a simplified sense, text-generating models try to predict the next most suitable word in a sentence, which makes models predictable. The more data was used to train the model, the better the AI can guess the next word. Some automated detection tools attempt to identify whether the sequence of words in a sentence is statistically optimal, indicating that the text was written by AI. Meanwhile, other tools engage in the reverse process, calculating the probability that the text was written by a human by analysing human writing structures.



Below are a few examples of automated tools. The tools have a field where you can paste suspicious text. After clicking the analysis button, the tool will usually provide a probability score indicating generated by AI. Some tools will highlight areas in the text that are likely generated rather than written by a human.

**Copyleaks**  

Examples: [GPT4](#) [ChatGPT](#) [Bard](#) [Human](#) [AI + Human](#) Model:

**Dear User;**  
I trust this email finds you in good health. I am writing to bring your attention to the importance of completing the registration/login process on our platform.  
As part of our commitment to providing you with a seamless experience, we kindly request you to finalize your registration or log in to your account at your earliest convenience. This will not only ensure the security of your account but also grant you access to the full range of features and benefits available on our platform.  
To complete the registration/login process, please follow the link provided below: Login Link  
If you encounter any difficulties or have any questions, please do not hesitate to contact our support team.  
Thank you for your cooperation, and we look forward to serving you.

[Clear](#)

**AI Content Detected**  

Zerogpt



### Your Text is AI/GPT Generated



Dear User!

I trust this email finds you in good health. I am writing to bring your attention to the importance of completing the registration/login process on our platform.

As part of our commitment to providing you with a seamless experience, we kindly request you to finalize your registration or log in to your account at your earliest convenience. This will not only ensure the security of your account but also grant you access to the full range of features and benefits available on our platform.

To complete the registration/login process, please follow the link provided below: Login Link

If you encounter any difficulties or have any questions, please do not hesitate to contact our support team.

Thank you for your cooperation, and we look forward to serving you.

■ Highlighted text is suspected to be most likely generated by AI\*

774 Characters

129 Words

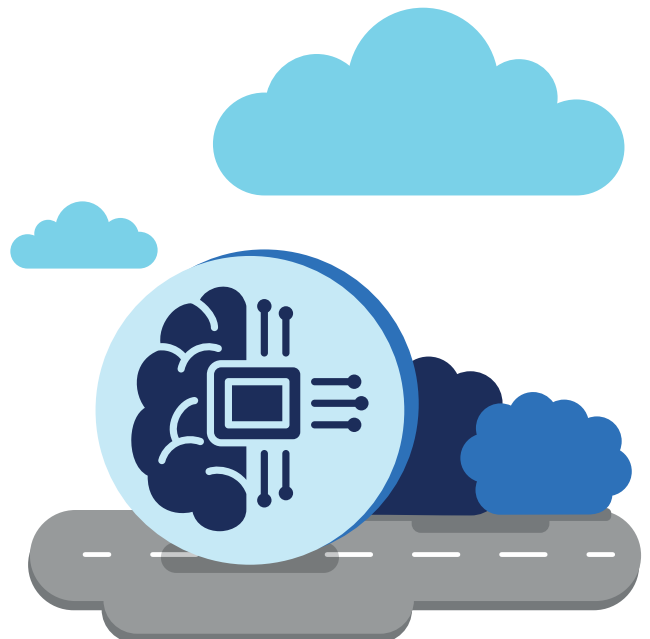
Writer



Gptzero



llm



## Information (text) checking

# How to check if text information is genuine?

### Google, Google, Google!

Search engines like Google are your best friend in the fight against misinformation. They help you check all the information in your browser to see if it is true or false.

Use keywords in your search that are based on the information you are interested in, so you can find out if the news you are interested in is being reported by credible sources. If you see suspicious information on social networks, there is a good chance that someone is already working to expose the false information and reveal the truth.

We have a big advantage in the fight against misinformation if we can successfully identify keywords. We can also check non-digital information and thus evaluate the information coming to us from different sources. It is important to stress that using a search engine will allow you to find a trustworthy source with the same content that you are trying to check, but first make sure that the source is credible.

### How to check?

1. Identify keywords best describing the piece of information you are looking into.
2. Use one of the available search engines to look for the same information.
3. Choose to filter your sources by „News“ and the most recent posts published in the last hour or day.
4. Identify reliable sources, and verify the information.

### Main rules to remember:

- Googling is the best first step to check information. It is the first step in the search.
- No matter what information you are interested in checking, Google searches can work very well: with text, photos or videos. A Google search can be very efficient.
- The most important thing is to use keywords so you can find the information you are looking for in a reliable source.

## To make more effective use of Google Search:

1. Use basic Google Dorking tricks, i.e. if you search for a phrase, put it between quotation marks ("..." or "...").
  2. If you have a keyword that is very popular in your search results, put a minus sign (-) next to it so that it doesn't appear in the search results.
  3. If you don't know the exact spelling of a word or the exact number or date, you can add an „\*“ as a wildcard to replace a missing letter in a word, a whole word, a number or a date.
- If you speak foreign languages, you can use this. Look for what sources in neighbouring countries or across the region are saying about the news you are interested in.
  - If you only speak one language well, get help from someone you trust who can check the relevant information in other languages. Later, discuss how the information of interest is presented in other language spaces. Such practice can enrich both your knowledge and that of the facilitator.

## Useful tools:

Google



Bing



Yandex



*(IMPORTANT: Be careful, this is a Russian site, so use extra security measures on your computer. As far as we know, the website is safe to use, but we recommend that you do not use the mobile app - it asks for access to personal data when installed.)*



## Image verification

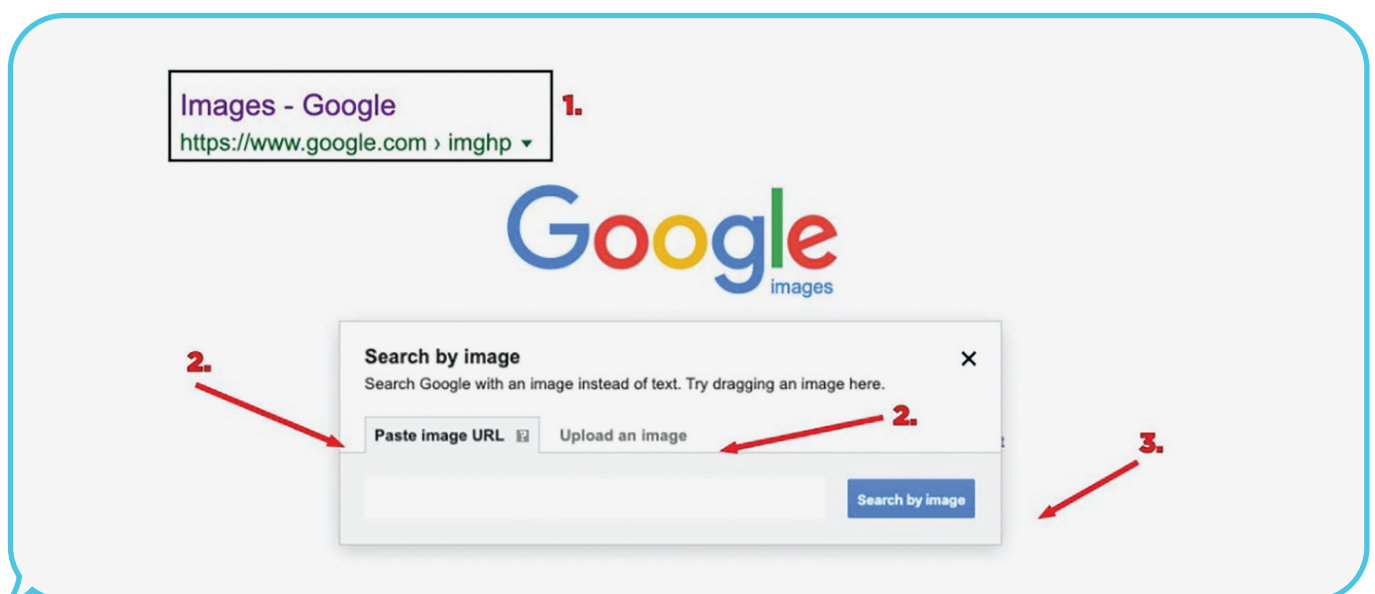
# How to check if visual information is genuine?

### Reverse Image Search:

Image recycling - posting a previous image and claiming that it was captured recently - remains one of the main problems in the disinformation space. When it comes to fake or modified images, the best technique to identify and verify them is reverse image search. This allows us to find all previously published images that are identical or very similar. Finding that an image has been published before is a reliable way to confirm that the image has been on the Internet for a long time. In other cases, if the image that caused the suspicion has been altered, reverse image search can help find the original image.

### How to check?

1. Open one of the search engines (links are provided next to the „Useful tools“).
2. Copy in the link or the downloaded image itself.
3. Investigate if the same or very similar images were posted before.



## Error level analysis.

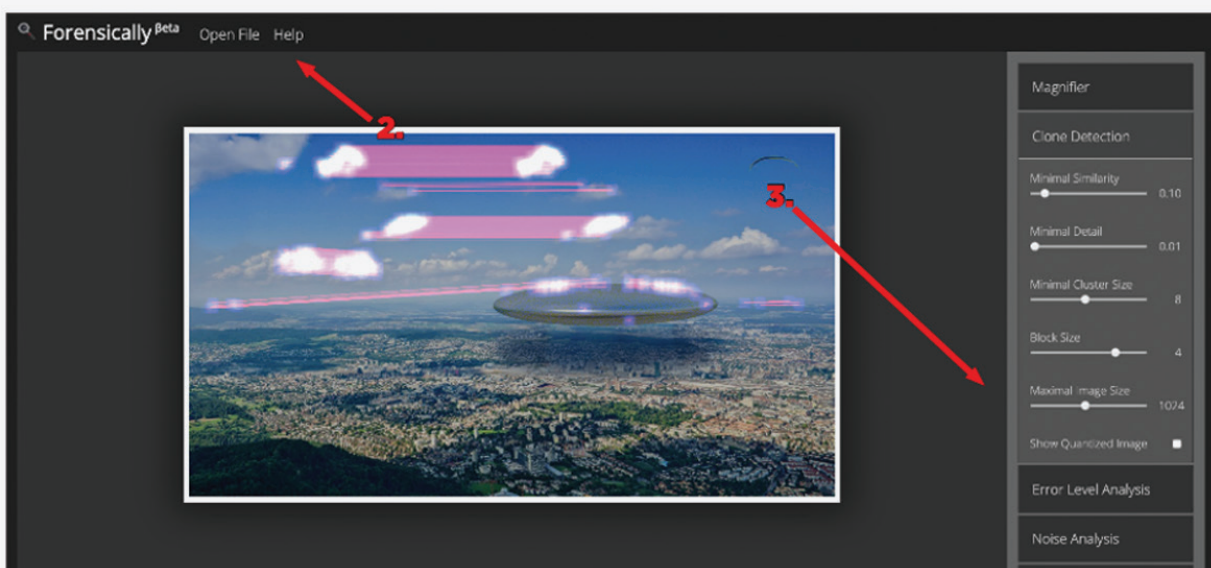
Error level analysis (ELA) is a more advanced method that permits identifying areas within an image that are at different compression levels. With JPEG images, the entire picture should be at roughly the same level. If a section of the image is at a significantly different error level, then it likely indicates a digital modification. In practice, you should look around the picture and identify the different high-contrast edges, low-contrast edges, surfaces, and textures. Compare those areas with the ELA results. If there are significant differences, then it identifies suspicious areas that may have been digitally altered.

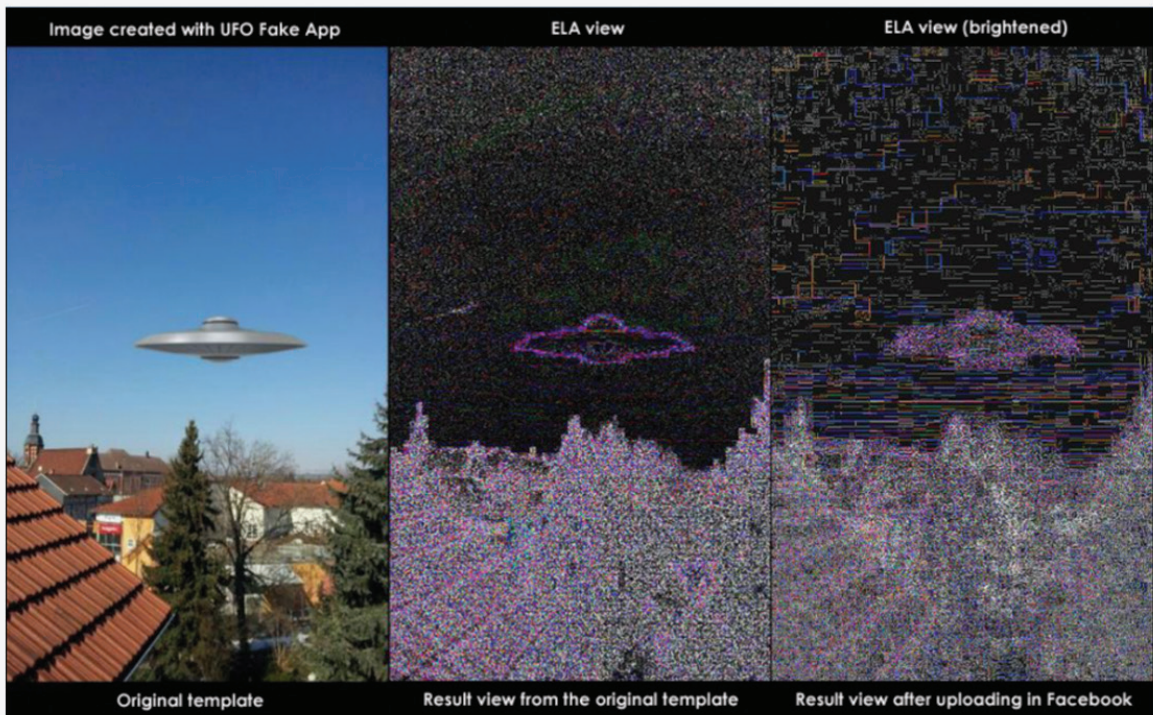
It is important to note that while this method is not bulletproof, it is nonetheless a reliable first step identification of digital alterations. Photo forensics is a separate scientific branch and to debunk expertly photoshopped images requires advanced knowledge and skills. However, in terms of everyday propaganda messages, the images are not well designed and are easily identifiable.

## Main rules to remember:

- Reverse image search should be a standard practice before trusting an image. If the image will not be what it said it was, it could deflect the actual genuine message;
- ELA is not a bullet-proof method, but it is a great quick way to check if the image is photoshopped.
- Reverse image search can be used to identify unknown people in the image.

Forensically, free online photo forensics tools - 29a.ch  
<https://29a.ch/photo-forensics>





**Useful tools:**

**Google** / Reverse Image Search



**Yandex** / Reverse Image Search



**Google** / RevEye extension



**Forensically**



**Foto Forensics**





# Generative Artificial Intelligence (AI)

## How to identify AI generated images:

Identifying a generated image, unlike generated text, is easier, especially when attempting to generate realistic photos. A few examples are provided below. The photos depict people in the old town of Vilnius. At first glance, it seems that the photos are real, but let's take a closer look:





- Colours – in the photo on the right, the colours appear unnatural and overly bright, straining the eye.
- Background – often, generated photos have an unclear background, or the lines of buildings or cars in the background are frequently distorted.
- Anomalies and distortions – upon zooming in on the photos, you will immediately notice that certain body

parts appear unnatural and distorted: ears look peculiar, heads seem as if painted with a brush, and so on. Additionally, AI often adds more than 5 fingers or more teeth than a human would have. It is essential to pay attention to the details.


- Watermarks – tools, especially free ones, often add their watermarks to photos. In the examples, you can notice the coloured squares in the bottom right corner, indicating that the images were generated with the help of DALL-E. Of course, we can manually remove this watermark. However, platforms, which let you generate images, started to add watermarks that are not visible to the naked eye. If the photo is uploaded to a verification tool, it will immediately detect that it is a generated image because it will identify hidden watermark.



Several automated tools can help identify generated photos. The principle of using these tools is standard. Simply drag and drop or upload the photo onto a specific section of the page and click “analyse”. After a few seconds, you will receive the result:

**Hivemoderation**  

Upload images here to test our model in real-time!  
Supports png, jpeg, jpg, webp. Use is subject to this site's [Terms of Service](#)



Upload

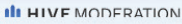
RESULT

The input is: likely to be AI Generated

99.9%

BY CLASSES

Classes	Score
<span style="color: #e91e63;">■</span> ai_generated	0.99
<span style="color: #00a0e3;">■</span> dalle	0.99
<span style="color: #e91e63;">■</span> not_ai_generated	0.00
<span style="color: #00a0e3;">■</span> none	0.00
<span style="color: #00a0e3;">■</span> midjourney	0.00
<span style="color: #00a0e3;">■</span> stablediffusion	0.00

 HIVE MODERATION

**Aiornot**  

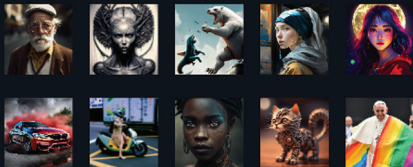
## Try AI or Not


IMAGES

AUDIO

### AI or Not

Determine whether an image has been generated by artificial intelligence or a human





Drag and drop

or upload your image

We support jpeg, png, webp, gif, tiff, bmp.  
10Mb of maximum size.  
User usage

OR

s
AI OR NOT?

## Video verification

# How to check if the video is genuine?

### Reverse image search.

Similarly to verifying images, the best methodology to check if a video is genuine is to reverse image search. Since videos are just a series of images, taking out a frame and searching for it is a great way to do it. Both tools InVid and Amnesty Data Viewer will allow you to find similar or identical videos already posted online, by looking for both frames and thumbnails.

### How to check?

1. Open one of the search engines (Amnesty DataViewer or InVid).
2. Insert the video link of the video.
3. Check if the video appears among duplicates.



## Youtube DataViewer

XI EPICdR + COLPIN / Corrupción judicial / Elber Gutiérrez

Video ID: Neo2Rp87Ifs  
Upload Date (YYYY/MM/DD): 2018-11-13  
Upload Time (UTC): 18:28:16 (convert to local time)

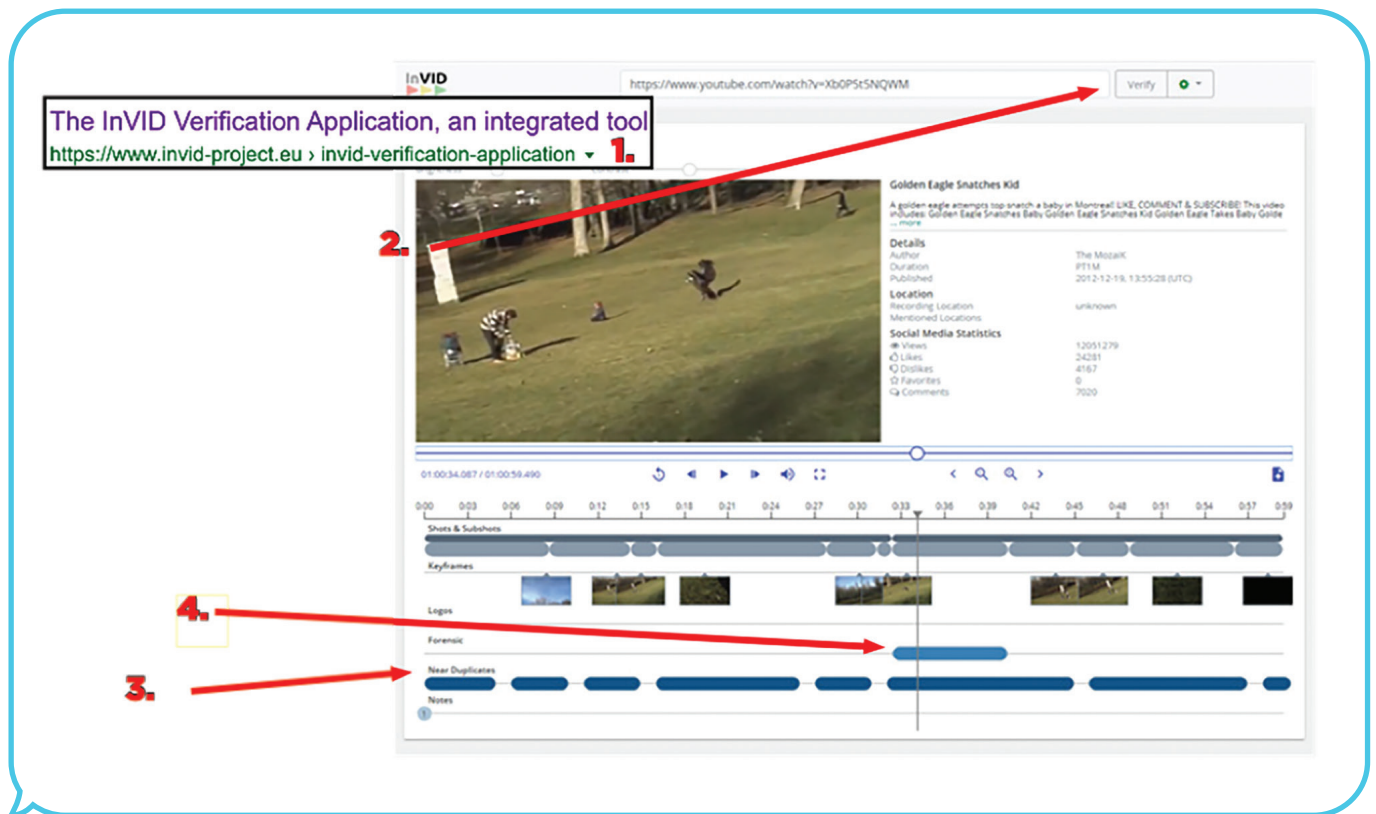
### Thumbnails:



reverse image search

## Forensic analysis

InVid also has the option of forensic analysis, when the software identifies potentially altered frames. The potentially altered frames pop-up in the analysis window, next to the sign 'Forensic'. If frames are identified by InVid as potentially altered, the chances are high that the video is fake.



## Main rules to remember:

- The main threat with videos is the same as images- video recycling. Previously produced videos are being re-posted and presented as a piece of fake news.

- These tools are not as effective as image verification tools, yet they can capture a lot of the fake videos.

**Useful tools:**

**InVid**

**Amnesty International YouTube Viewer**

## Trolls

# How to spot a troll online?

## What is a troll?

A troll is a person who intentionally initiates online conflicts or offends other users to distract and sow divisions by posting inflammatory or off-topic posts in an online community or a social network. Their goal is to provoke others into an emotional response and derail discussions. A troll is different from a bot because a troll is a real user, whereas bots are automated. The two types of accounts are mutually exclusive.

To spot a troll is harder than to spot a bot, as these accounts are usually more sophisticated and are actively pretending to be real people. Below you can find a number of criteria that will help you to identify a troll, but these clues are indicative, rather than conclusive.

It is seldom possible to say with 100% certainty

that a given account belongs to a troll operation rather than merely supporting certain malign narratives. Before examining the factors which reliably indicate a pro-Kremlin troll, it is important to look at one factor which does not - hyper-partisan content. A variety of contemporary real social media users tend to be highly partisan, especially when it comes to political topics.

The following are some criteria to help you identify trolls but be warned that these hints are indicative and not definitive. It is rarely possible to be 100% certain that an account belongs to a troll and not to a negative user.

---

### 1. Mistakes in articles: the English articles "a" and "the"

One of the linguistic signs which is characteristic of many known Russian accounts is the inability to use the grammatical articles — "a" and "the" — appropriately. The Russian language has neither.

---

### 2. Mistakes in formulating a question

Another common linguistic indicator is the inability to phrase a question. In Russian, the word order for questions does not change, unlike in English, German, and formal French. Many known Russian troll accounts have posted questions which kept the word order of statements.

---

---

### 3. Unclear or questionable identity

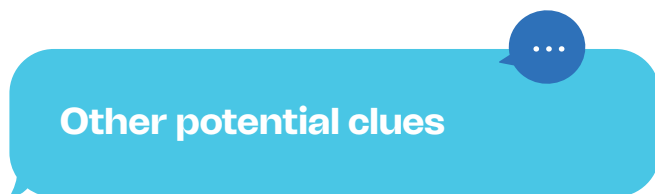
Some trolls are using fake names that are very common in a given language, making it difficult to differentiate the specific author or deliberately leading to mistake it with another, such as a recognized journalist. The names used by trolls are also intended to be perceived as traditional or "sound right", so that any reader would tend to trust or not question the authenticity of an author of an article or a comment on social media.

It may also be useful to check their profile pictures if there are any. Such pictures, added for achieving additional trust, may be stock photos that you may easily find on the internet. Such pictures may also be deliberately unclear when looking closer (photo editing, supposed person wearing sunglasses, etc.), making it impossible to clearly identify the person.

---

### 4. Amplification of pro-Kremlin narratives

The Russian government has developed a distinctive narrative on key geopolitical events of the last five years. This follows the principles established as early as in the Doctrine of Information Security of the Russian Federation (2000) on conveying state policy and official positions on issues that are important to the Russian government. As pro-Kremlin narratives are widely available on online sources, such as the Russian Ministry of Foreign Affairs or the RT Twitter account, it is easy to check if the same themes appear in the suspected account. An account which repeatedly shares Russian government talking points on most or all of these events can justifiably be considered pro-Kremlin.



---

#### Trolls have throwaway email addresses:

Since many pages or blogs that allow comments on articles also require an email address, trolls get around this by entering fictitious ones. Most disposable email addresses are random and easily overlooked because they do not reflect the person's real name.

---

#### The aim of trolls is to cause people to be outraged:

They are not polite and are not ashamed to engage in open warfare. They hurl accusations and usually sound angry.

---

#### Trolls use anonymous proxies:

Trolls often use anonymous applications or proxy servers that display a different Internet Protocol (IP) address than you are used to.

---

#### Trolls rarely say anything of value throughout the conversation:

When trolls enter a community discussion, they do not add anything meaningful to it. Instead, they make jokes, reproaches and insults.

---

## Fake Facebook accounts

# How to spot if the Facebook account is fake?

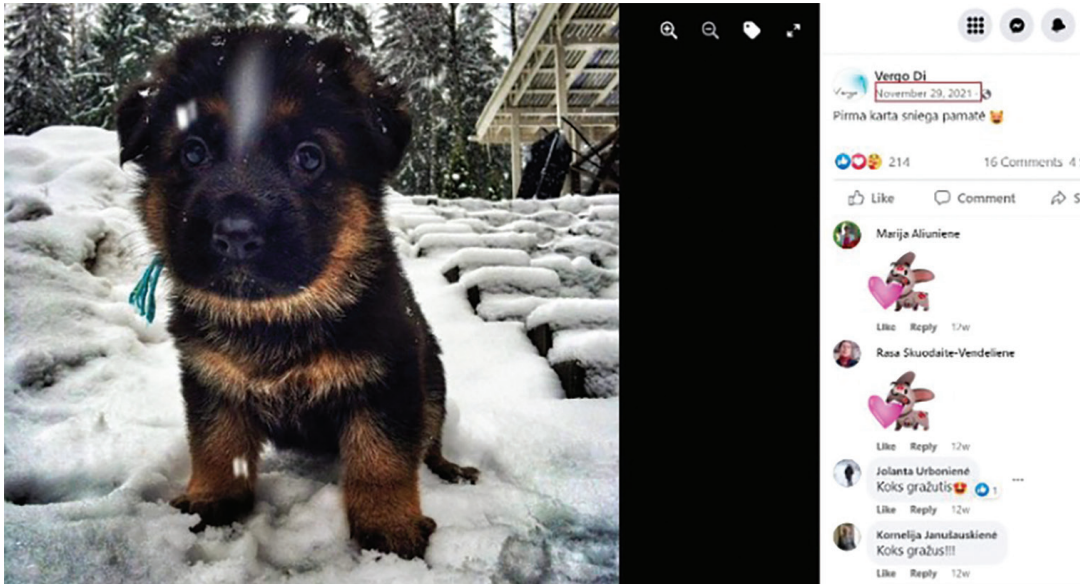
Fake social media accounts are not as active as trolls and usually play the role of silent spectators. Similar criteria for fake accounts are applied on most platforms, but we have chosen Facebook as the main example here. Real Facebook users often share personal information on their accounts, so these accounts actively try to become your friend for two main reasons:

to appear more real because they have a lot of real people on their friends list, and to be friends in order to see more personal information. Depending on the purpose of the fake account, it can be used to collect personal information from employees of various organisations.



## 1. The factor of attractiveness

Accounts of visually appealing users you don't know who invite you to become their friend may be fake.



We can easily check that it is not an image copied from the internet, chosen by the owner of the fake account to attract your attention and friend you.



### 1 result

Searched over 52.5 billion images in 0.5 seconds for:  
[lh3.googleusercontent.com/J3zMaRV\\_\\_qxt5dSuOpufIhxiMfP39e...](https://lh3.googleusercontent.com/J3zMaRV__qxt5dSuOpufIhxiMfP39e...)

Sort by best match ▾

Filter by website / collection



### instagom.com

[explore/tags/germanshepherdpuppies](https://explore/tags/germanshepherdpuppies) - First found on May 28, 2017

Filename: [18722090\\_1019473188187238\\_8466430548250722304\\_n.jpg](#)  
(960 x 960, 151.9 kB)

At <https://tineye.com/> the image search shows the date when the photo was first shared. We can conclude that the profile is a fake, audience-attracting questionnaire, sharing fake posts and occasionally propaganda messages.



---

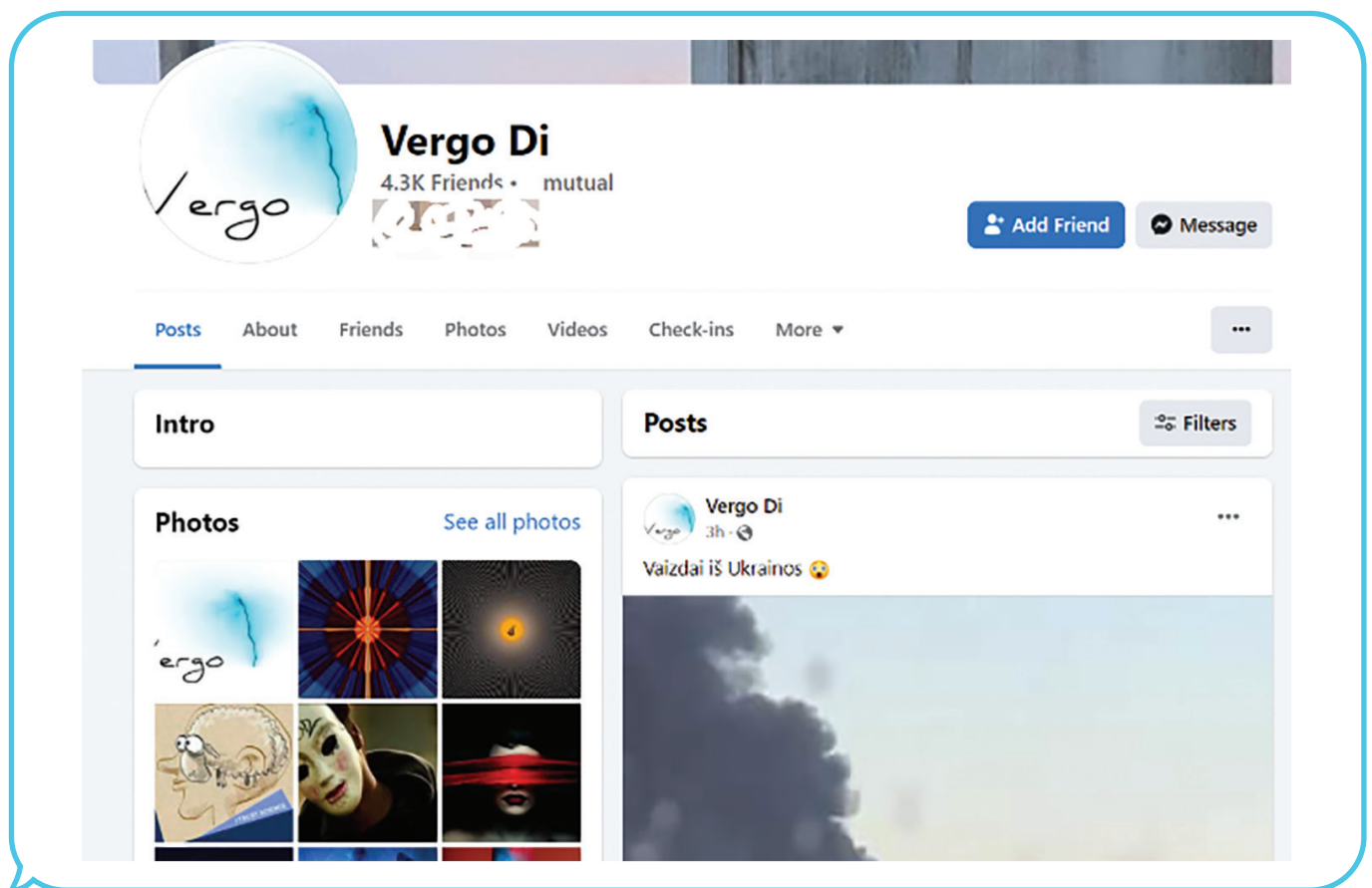
## 2. Few photo uploads

Most fake accounts don't post many photos - usually three or four, sometimes of different people. That's enough photos to create the temporary illusion that there is a real person behind the account.

---

## 3. Strange biographies

Most fake accounts have biographies that contain very little information or seem strange. For example, it is not impossible, but highly unlikely, that a person living in the Bronx went to the University of Helsinki, is also very young, and is already working for a New York PR firm. A quick check of his name in a Google search and a reverse search of his profile picture can show that the account is fake.



---

## 4. Unresponsiveness

If you send a message to a fake account, it is unlikely that you will get a reply to even a short question. It is best not to even try to contact or otherwise engage in conversation.

---

## 5. A mostly blank Facebook wall

The only things you'll find on one of these fake Facebook account walls are new „likes“ on Facebook business or product pages and new friends.

---

# Response

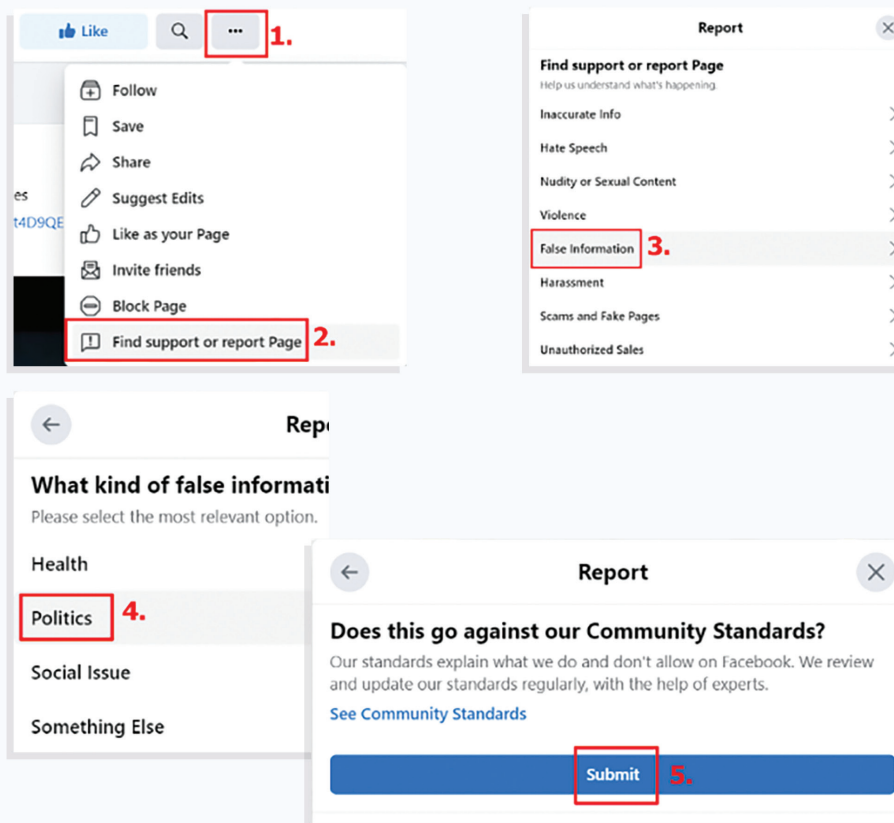
Proactively tackling disinformation on the internet requires two steps: exposing it and reporting it.

## Facebook / X / media reports



It is very important that your friends, classmates or colleagues are aware of the disinformation being spread against your organisation. Every organisation must have clear procedures in place to ensure that its members know where to send a report of a false story they have seen. The main purpose of this is to inform the members of the organisation in a unified way that a particular message being spread is false and to prevent them from sharing and believing the story.

The second step is to report it on a social networking platform. All social networking platforms have a function to report a news item or post in any format, with a specific reason for why it is being reported. If the social networking platform receives enough user notifications, the created story or fake post will be deleted. This is the method used by civil society organisations to combat online disinformation. If media outlets are spreading fake stories, depending on the nature of the media outlet, this should be reported either to the media outlet itself or to the national media control authorities. For example, an e-complaint can be sent to the Inspector of Journalistic Ethics: [https://www.zeit.lt/lt/e.skundas/560?fbclid=I-wAR3d1M6DGEEnoolxRLHLqGLP7v\\_oNVb\\_tEILC-2WhzMul2piC7WE46dwYtj-](https://www.zeit.lt/lt/e.skundas/560?fbclid=I-wAR3d1M6DGEEnoolxRLHLqGLP7v_oNVb_tEILC-2WhzMul2piC7WE46dwYtj-)



## Risks and dangers of using "Telegram"

As many young people use Telegram for their daily correspondence, they are also aware of the platform's other features - the groups and channels that can be private and anonymous. While open communication flows are common in social media, closed communities can disseminate specific content, including biased and distorted information, whose origin is very difficult to verify.

Research shows that Telegram channels and groups have been used in some disinformation campaigns to spread hoaxes. For example, the messages disseminated in various languages included information on the COVID-19 pandemic, as well as pro-Kremlin views on the war against Ukraine, far-right rhetoric, and conspiracy theories.

„Telegram channels and groups can also be used to mobilise users for ideological rallies or political protests. If these events are organised democratically and transparently, there is no question or problem, but sometimes the real beneficiaries are hidden and there is little or no information about these real persons. For example, several anonymous Telegram groups and channels were used to disseminate provocative calls during the „I'm Russian" campaign in Estonia:

<https://eng.obozrevatel.com/section-life/news-russians-in-tallinn-threw-a-hysterical-tantrum-because-the-police-forced-them-to-remove-i-am-russian-stickers-from-their-cars-video-27-09-2023.html>



As many disinformation disseminators and propagandists (including pro-Kremlin ones) use Telegram, it is easy to share their content with other groups and communities led by anonymous administrators. It is one of the most widespread channels for the digital dissemination of anti-Western or anti-liberal disinformation. For example, the Telegram channel „Antifашисты Прибалтики” (Antifascist Baltic) has been responsible for inflating the number of views of the „Russophobia” narrative, with its posts garnering as many as a few hundred thousand views each.

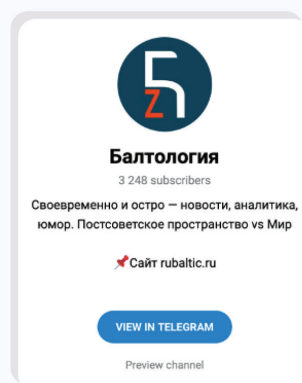
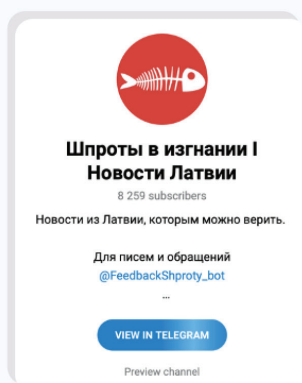
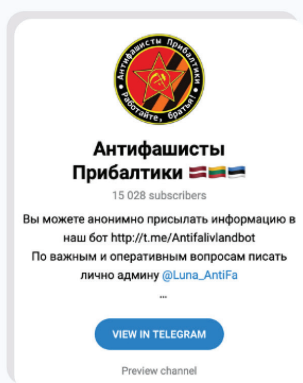
One of the posts on the channel talked about Latvia’s alleged Russophobia because a puppet theatre decided to ban the performance of “cheburashka” (a character from Soviet cartoons). The post also incited hatred against the Latvian Minister of Culture by uploading a (fake) photo of him posing in front of grotesque objects, claiming that “this is the face of Latvian culture and national identity”.

Another entry stated that any “free West” person who dared to even mention Russia’s right to defend the Russians would be immediately put behind bars, with their property confiscated and banned from all economic and creative activities. These records show that Russia has actively tried to spread the narrative that Russian culture is under attack in the Baltic States and that the Russian-speaking minority in the countries is unable to express itself for fear of repression.



**A few simple tips that can help you strengthen your digital hygiene on your Telegram channel:**

- Before joining a group or channel, make sure that the content is 100% of your interest, and search and ask for more information about the online community administrator.
- Be vigilant - anonymous groups and channels are extremely likely to contain unverified and/or obscure information if the topic is relevant and important. Content must have credible sources and must not contain speculative opinions, „alternative facts” or simplistic propaganda clichés.
- If you are emotionally provoked by any information in a Telegram group or channel, ask yourself why this has happened and why it is useful - do not be quick to react or share information that may be biased, polarising, offensive or simply false.
- Any suspicious posting in a Telegram group or channel can be reported to administrators, the web police, and the digital fact-checking community (e.g. CRI in Lithuania or Propastop in Estonia). Before doing so, make sure that you keep as much of the original recordings as possible, such as a screenshot with textual or visual information.





## Risks and dangers of using TikTok

TikTok, which has recently gained popularity among young people, is of Chinese origin and is characterised by a high level of online misinformation. The creators of the app themselves say they are working hard to combat misinformation, radical extremism and hateful behaviour, but how are they really doing?

Although the social network did not seem dangerous at first, as the number of users grew over time, the content began to change. Now there is no shortage of posts on the network that spread pro-Kremlin **narratives\***.

**\*A narrative** is a systematic and coherent story that is created by repeating messages on a certain topic, adding new facts and context to these messages.

**A narrative is a story that convincingly conveys a key message, a formation of opinion.**

The TikTok platform is characterised by very vague algorithms and respect for the laws of authoritarian regimes. This social network is also distinguished from others by its addictive effect. It is manifested by the constant viewing of short videos that are rich in emotional elements and have a catchy soundtrack. The more time spent on TikTok, the better the algorithmic presentation of information becomes, with propaganda messages appearing in the overall stream of content consumed.

**There are two main challenges to using TikTok:**

• **Aggressive data collection.** When the app is installed on a phone or other smart device, it asks for

more access to data and then collects it. The risk of using TikTok is that the app can see the user's contacts, as well as which other apps are being used on the device, to find out more details about the location and to identify where a particular device is. There is also a risk that correspondence that takes place within the app may also be tracked by certain keywords and come under the company's radar.

• **Shadow banning.** If a user posts something that TikTok developers do not like, the post may be hidden, i.e. a shadow ban will be enforced.

TikTok uses the algorithm as a tool to attract and retain attention, and it strives to provide a personalised experience for each user. The algorithm uses data collected from users to determine what content might be of interest to them. For example, the longer you watch a video, the more likely you will see similar videos on TikTok in the future. The app also remembers your search keywords so it can suggest videos of the same style. In addition, TikTok's algorithm connects users with common interests by showing them similar content. If you see the same videos as some of your friends, it's no coincidence.

Remember that TikTok was not designed to deliver news - short videos are more tailored to the user's entertainment. Scrolling and watching videos on TikTok activates the parts of the brain responsible for the feeling of success - just like gambling in the hope of lifting your spirits, but instead of real gains, it wastes a lot of time. Because TikTok videos are short, entertaining and easy to access, it creates a kind of addiction that leads to a loss of focus and timing.

TikTok also exploits your curiosity and fear of missing out on something important. Have you heard the saying "If you're not on TikTok, you don't exist at all"? It's called manipulation, and it's designed to force young people to keep continually using the app.

Another trick used by TikTok is anger management. The app offers videos that offend a group of people, an ideology, or a movement in the hope that such content will offend viewers and start an emotionally heated debate, which attracts more attention and the videos go viral. These techniques are used to increase the number of clicks and followers. TikTok is plagued by cyberbullying incidents, and users are therefore at high risk of becoming a victim of harassment or hate speech. In addition, TikTok content can be created by anyone, so it is bound to contain biased or manipulated information. The large amount of personalised content on TikTok makes it even harder for young users to distinguish between the opinions of certain users and the facts.

**TikTok can be a usable social network if you follow the simple tips for using it:**

- Limit your daily time spent on TikTok - meeting friends in real life is always more fun than online.
- Remember to be critical of news-like content on TikTok - double-check the information you are interested in from other sources (not social media).
- Report cyberbullying and hate speech - do not distribute offensive videos or emotionally sensitive content.
- If you feel tricked or manipulated, talk to your parents or the e-police.

**Extended measures and recourse to assistance**

For more serious cases of disinformation, there are two main approaches: using more sophisticated open-source techniques or seeking help from the online research community. Most online tools are relatively easy to use and provide step-by-step instructions on how to do so. The following are two large and most useful tools:

**Bellingcat's Online Investigation Toolkit**



**Online Open Source Tool Box**



If there are still more questions than answers, please contact the online research community and provide them with information about the false story you have found. Most researchers will be happy to debunk a fake story and share the rebuttal content online.



Vilnius, Lithuania  
**Edition 500**  
©CRI, 2024



This publication is  
sponsored by

**Google**