

Platform Response to **Disinformation** during the EU Election 2024

Analysis of the responses of very large online platforms to debunked disinformation during the EU Election 2024



**FUNDACIÓN
MALDITA.ES**
PERIODISMO PARA QUE NO TE LA CUELEN

1. Executive summary.....	2
2. Platform response rates to election-related disinformation.....	3
3. Platform response rates by key election-related topics.....	6
4. Engagement with election-related disinformation.....	9
5. Methodology.....	15

1. Executive summary

The goal of this report is to evaluate how five very large online platforms (VLOPs: Facebook, Instagram, TikTok, X and YouTube) have responded to debunked electoral disinformation about the European Parliamentary elections of June 2024, in line with their obligations under the Digital Services Act (DSA) to “*put in place reasonable, proportionate and effective mitigation measures*” against disinformation.

For that purpose, we have analyzed the disinformation content identified by the independent European fact-checking organisations involved in the EFCSN's [Elections24Check](#) project during the four months ahead of the elections, including ElectionDay (from 1 february to 6 June 2024). Elections24Check is a database that gathers and categorizes fact-checked information in Europe ahead of the 2024 European Elections. Over 40 EFCSN member contributed to the database. The content assessed in this report covers a total of 1321 posts identified in 26 European countries.

The main results of this analysis are the following:

- All in all, **45% of those disinformation contents received no visible action by the platforms**, but each platform's response rate varies, as did the particular content moderation decisions they made when they did act.
- **YouTube** took no visible action regarding 75% of disinformation contents and, even in the cases where it did, in 80% of those it just surfaced a generic information panel or labeled the source of the video as state media, offering no explanation of why the content itself was false. Some of those videos reached 500.000 views.
- **X** took no visible action in 70% of the cases and Community Notes were visible in only 15% of the posts already debunked by European independent fact-checkers. Among the 20 most viral debunked posts that received no action by the platforms, 18 were hosted in X with over 1,5 million views each.
- **TikTok** managed to take visible action on 40% of the posts containing disinformation, Instagram in 70%, and Facebook on 88%. While the majority of Facebook actions regarding disinformation were fact-checking labels that kept the original content online and focused on adding context (77%), TikTok's most common action against those contents was to remove them (32%).
- The 20 most viral posts or videos that received no action by the platforms hosting them accumulate **over 1,5 million views each**.
- Across the most common topics, the one in which platforms' response was the least successful was disinformation targeting **migrants** (57% of the cases received no action) closely followed by disinformation about the **integrity of the election** (56%). Both YouTube and TikTok had a 0% response rate for disinformation targeting migrants.

2. Platform response rates to election-related disinformation

When analyzing disinformation content identified by fact-checkers under the [Elections24Check](#) project, we see that platforms have shown different degrees of response to the content debunked by the fact-checkers. Overall, **43% of the disinformation content that affected the election received no visible action from the platforms.**

	Facebook	Instagram	TikTok	X	YouTube	TOTAL
ACTION	88,83%	70,73%	40,00%	29,33%	24,49%	56,85%
NO ACTION	11,17%	29,27%	60,00%	70,67%	75,51%	43,15%

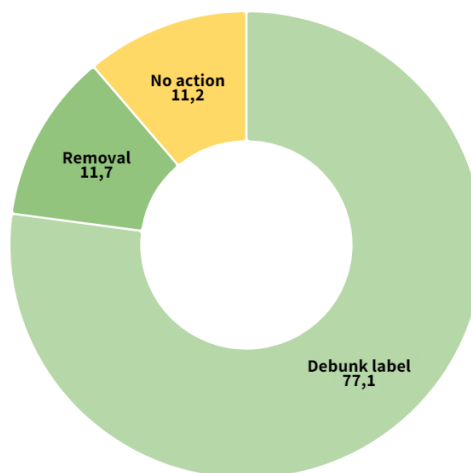
Per platform, only Meta's services have taken visible actions on most of the disinformation content, with Facebook taking action on 88% of the posts and Instagram on 70%. The rest of platforms have taken visible actions in less than half of the disinformation contents identified: TikTok (40%), X (29%) and YouTube (24%).

When analyzing the actions taken to address disinformation, we see how platforms prioritize different types of measures. In the case of fact-checking and debunk labeling, which has been proven to be the most [effective action to address disinformation while respecting freedom of expression](#), **Facebook (87% of actions) and Instagram (79% of actions) have been the only ones using it consistently and prioritizing it over other types of actions.** X has started using a type of labels (19% of actions) which are not supported by debunks (what we call here "generic label") and also in a less systematic way than Meta's debunk labels.

Facebook

Looking at the visible actions, Facebook has mostly used debunk labels, being the platform that has used it the most (in 77% of the content), while the rest of the actioned content was removed (11,7%). Although Facebook has been the platform with the least percentage of content without visible actions taken, still 11% of the content remains online without visible actions.

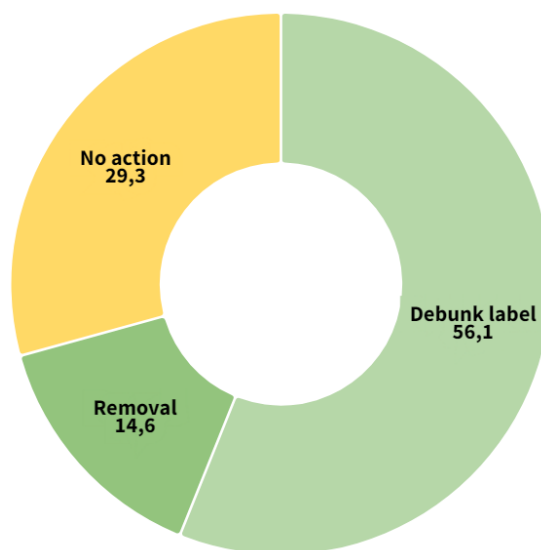
		Facebook
Removal		11,69%
Debunk label		77,14%
ACTION	TOTAL:	88,83%
NO ACTION	TOTAL:	11,17%



Instagram

Similarly to Facebook, but to a lower extent, in Instagram debunk labels are the predominant visible measure (in 56% of the disinformation content), with the rest of actions being removals (14%). While Instagram also belongs to Meta and shares its tools, the results are much less positive. It has also acted on most of the content but also on a lower proportion (70%) than in Facebook.

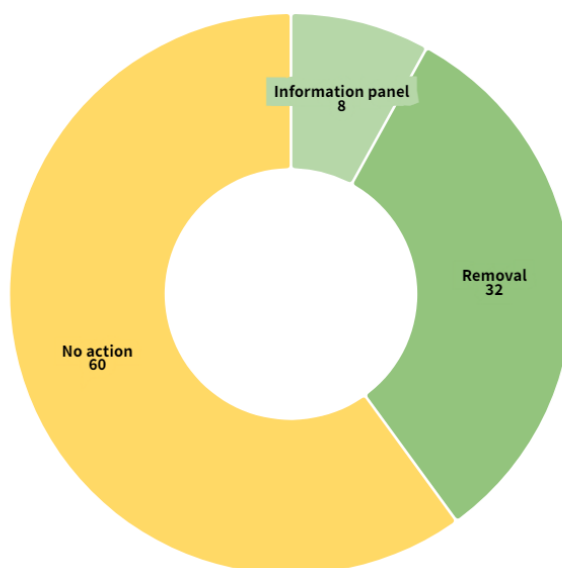
		Instagram
	Removal	14,63%
	Debunk label	56,10%
ACTION	TOTAL:	70,73%
NO ACTION	TOTAL:	29,27%



TikTok

TikTok has relied mostly on removals (in 32% of the disinformation content), while the rest of visible actions are information panels with a link to “info on the European elections” (in 8% of the disinformation content).

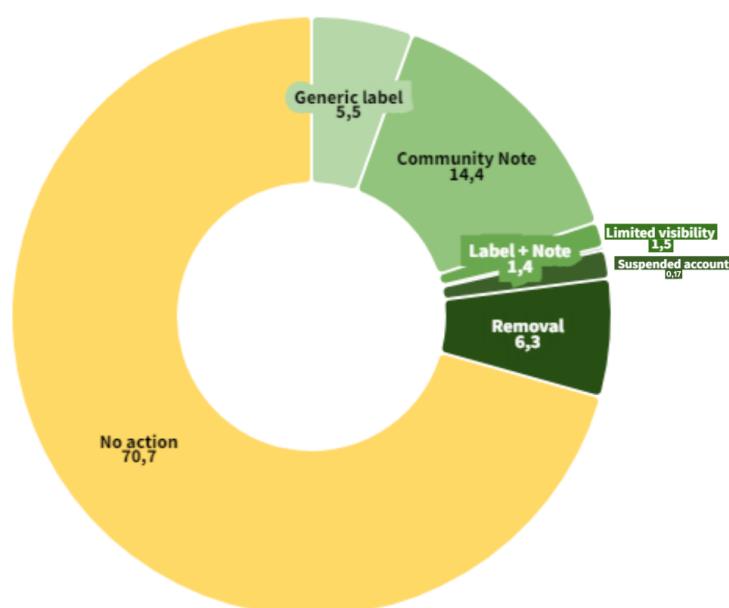
		TikTok
	Removal	32,00%
	Information panel	8,00%
ACTION	TOTAL:	40,00%
NO ACTION	TOTAL:	60,00%



X

The main visible action taken on X's disinformation content has been taken by the users (Community Notes in 14% of the disinformation content), followed removals (in 6% of the disinformation content), generic labels – tags on signaling manipulated or out of context content – (in 5,5% of the disinformation content), suspension of the account (in 1,54% of the disinformation content), combination of generic label and Community Note (in 1,3% of the disinformation content), and limitation of the visibility (in 0,2% of the disinformation content).

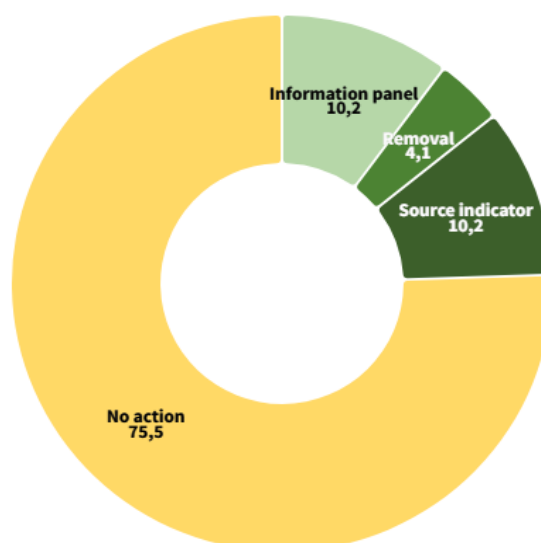
		X
	Removal	6,35%
	Suspended account	1,54%
	Generic label	5,49%
	Community Note	14,41%
	Label + Note	1,37%
	Limited visibility	0,17%
ACTION	TOTAL:	29,33%
NO ACTION	TOTAL:	70,67%



YouTube

YouTube is, by far, the platform that shows the least action on disinformation content, only on the 24,5% of the videos flagged. The main action taken has been the inclusion of information panels about climate change (in 10% of the disinformation content), same as source indicators regarding public media outlets (10%) and followed by removals (4%).

		YouTube
	Removal	4,08%
	Information panel	10,20%
	Source indicator	10,20%
ACTION	TOTAL:	24,49%
NO ACTION	TOTAL:	75,51%

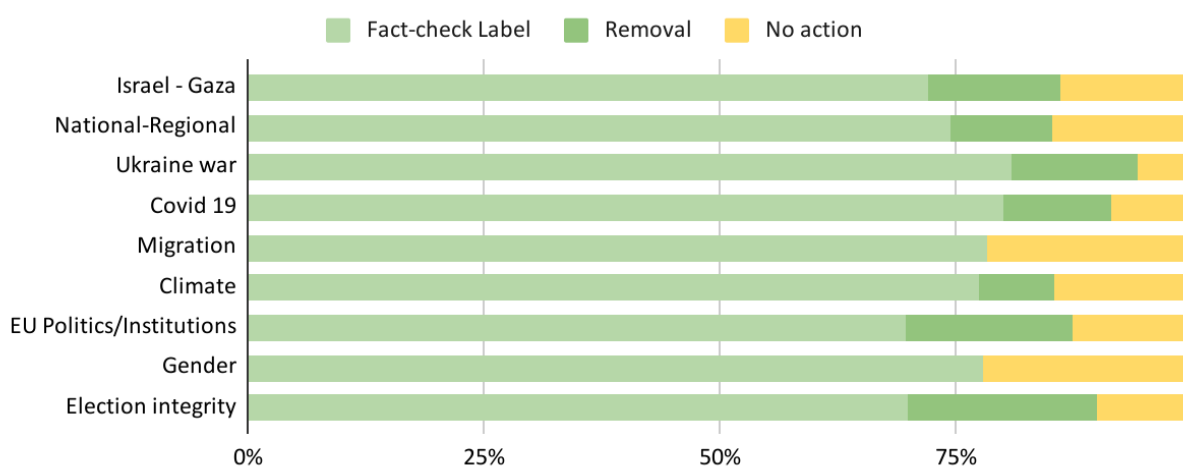


3. Platform response rates by key election-related topics

Disinformation related to certain issues experienced different levels of response by platforms. For this part of the analysis, we assessed the response rates for disinformation posts related to the most prevalent election-related topics, as well as those that are related to issues mentioned in the DSA as particularly relevant for systemic risk.

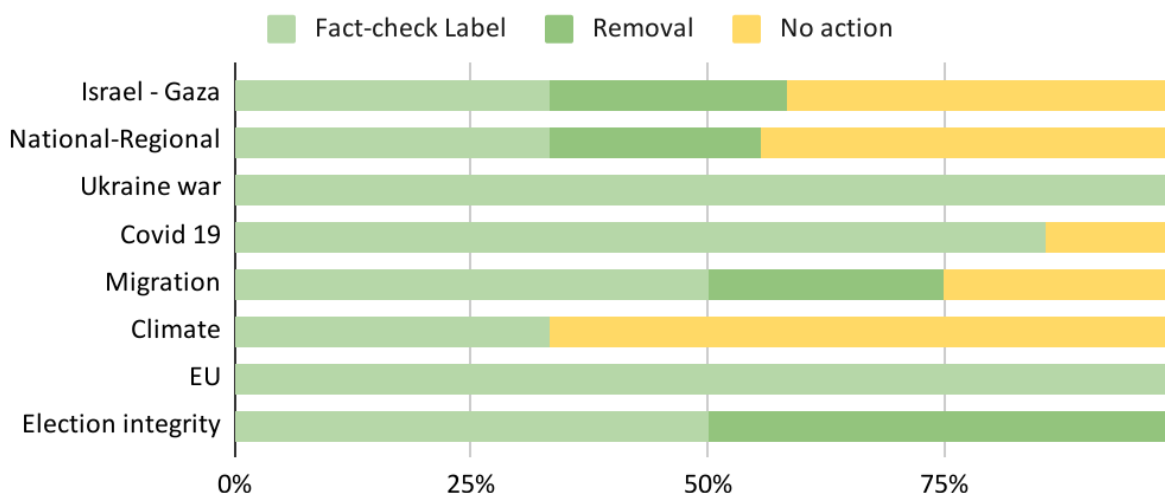
Facebook

- Moderation actions taken in Facebook posts were consistent: the rate of content actioned in every topic was above 75%



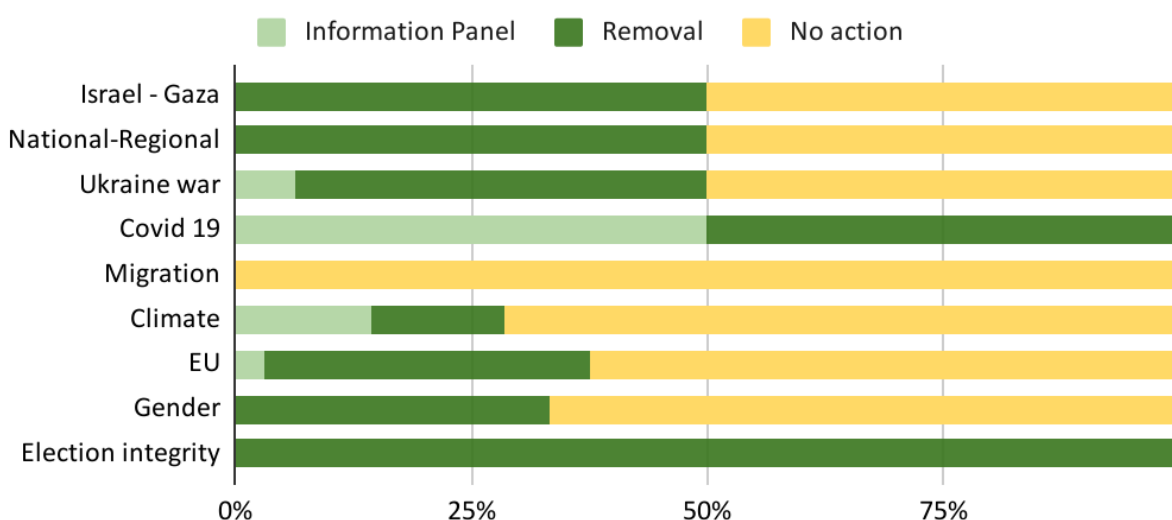
Instagram

- 2 out of 3 posts of related to climate disinformation did not have any visible action



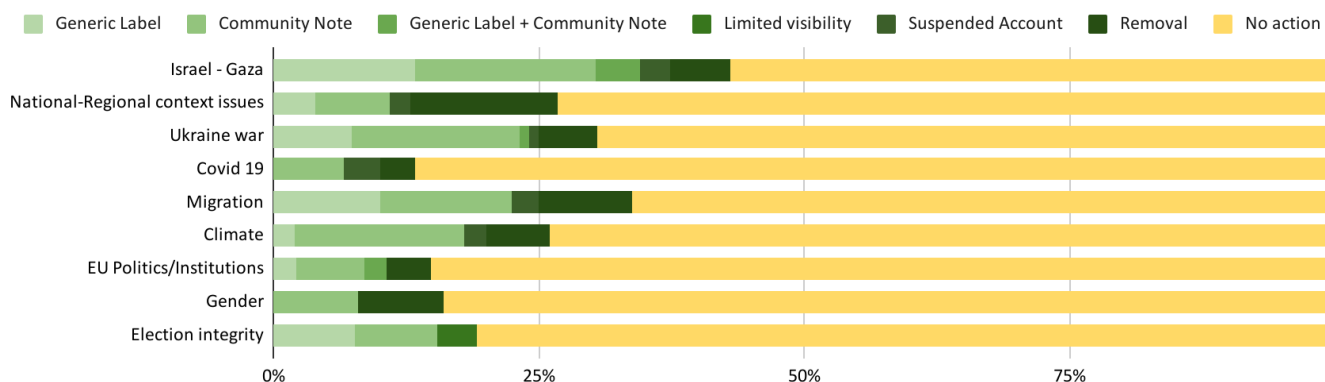
TikTok

- All disinformation content under the category of 'Migration' received no visible action
- More than two thirds of all disinformation about climate and gender went unactioned.



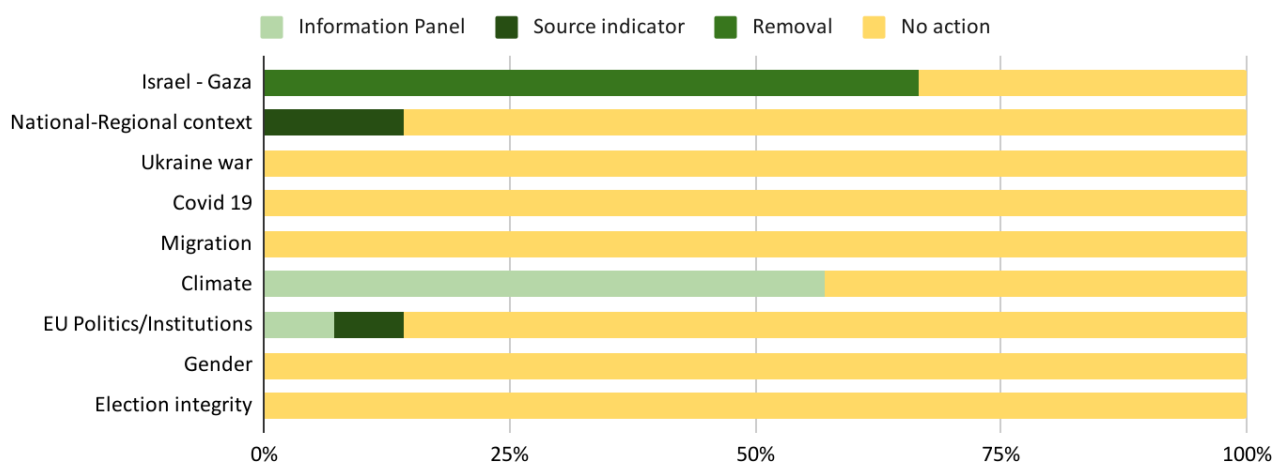
X

- X's response rate was the highest on disinformation content related to Israel and Gaza, but even on that, it still did not reach 50%
- Disinformation related to Covid-19 narratives had the lowest action rate (13%)



YouTube

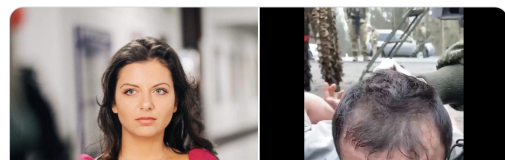
- All removal actions taken by YouTube on the disinformation content were on videos related to Israel - Gaza, possibly due to their overlap to other policies
- Content under five different thematic categories was not visibly moderated: Ukraine, Covid, Migration, Gender, and Election Integrity.



4. Engagement with election-related disinformation

Among the debunked posts that have not received any visible action by the platforms, we see many of them that are hugely viral. **The 20 most reproduced posts accumulate each one over 1,5 million views.**

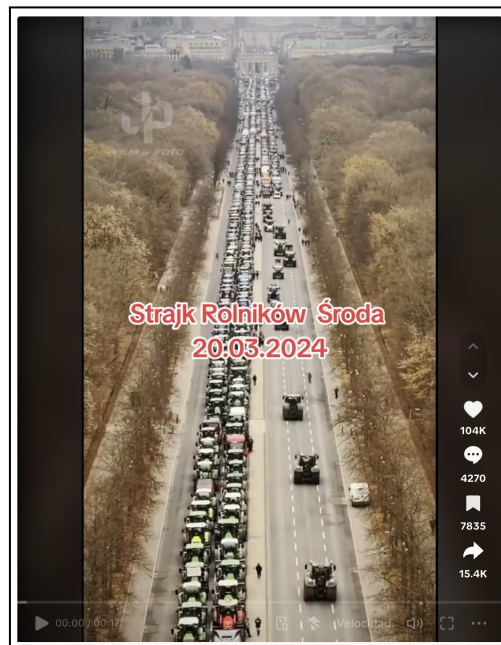
X is by far the platform with more viral debunked posts with no visible actions. **Among the 20 most viral debunked posts, 18 are hosted in X while the other 2 are in TikTok, all of them with over 1,5 million views each.** One of the most viral debunked posts that did not receive any visible action by the platform is a post affirming that “[Ukraine is behind the Moscow attacks](#)”. The tweet accumulates **over 3.4 million reproductions, 9,000 likes and 3,000 comments.**

 Victor vicktop55  @vicktop55 Margarita Simonyan: We already know the names of the performers. We have already seen their faces and know that they sing during interrogations. (We do not publish it so as not to interfere with the investigation). And immediately it becomes obvious why yesterday the American media were shouting in chorus that it was ISIS. Because it's not ISIS. It's just that the performers are chosen so that you can convince the dim-witted world community that ISIS. Elementary sleight of hand. The level of the station thimble. So that's it. This is not ISIS at all. These are Ukrainians.	And the fact that yesterday, even before the arrests, before the persons and surnames of the perpetrators, Western special services began to convince the population that this was ISIS - and set fire to the hat on this thief. This is not ISIS. This is a well-coordinated team of several other, also widely known, abbreviations. I am publishing the full version of the ghoul's interrogation. Half a million rubles, 'to shoot people of any kind.' t.me/vicktop55/2297... Traducir post 
---	--

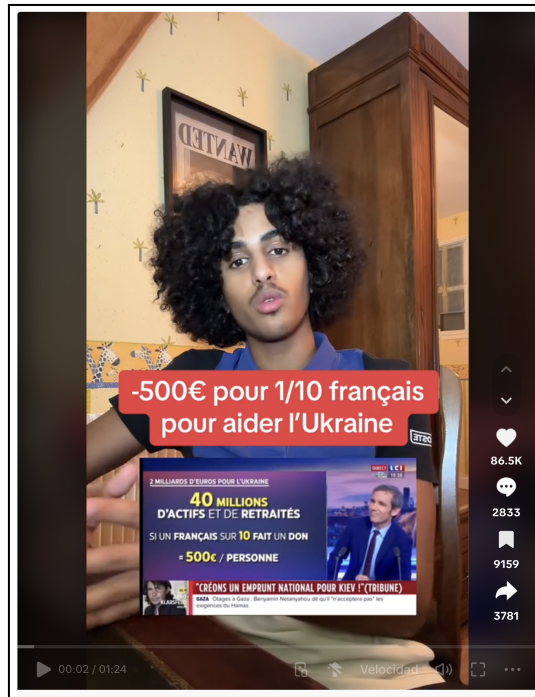
Another of the top-viral posts in X claims that “[Ukraine is sending teenagers to war](#)” with a video showing images of students from the military high school in Kiev, not soldiers being sent to the front. The video, shared by a Spanish account, accumulates **over 2,2 million reproductions, 3,000 likes and 2,000 comments.**



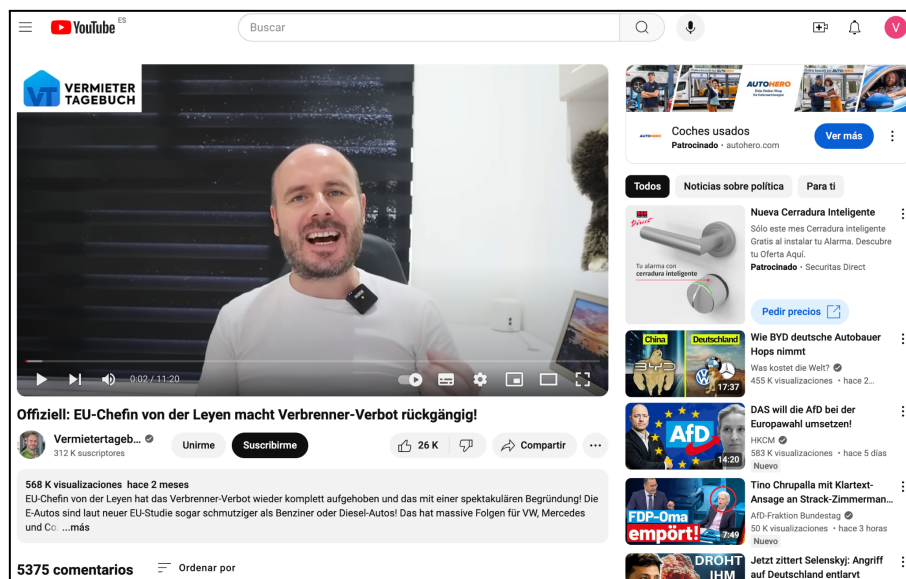
In TikTok, the most viral debunked post that did not receive any visible action from the platform is a video that is supposedly “[Depicting the farmers' protest in Poland](#)”, using a video from a protest in another country. The video had **over 4 million reproductions** and 100,000 likes and 4,200 comments.



Also in TikTok, one of the most viral videos with **over one million reproductions** claims that “[The \(French\) government will withdraw 500 euros from one in ten French people's savings account to help Ukraine](#)”. The video, published in France, has over 85,000 likes and 3,600 comments.

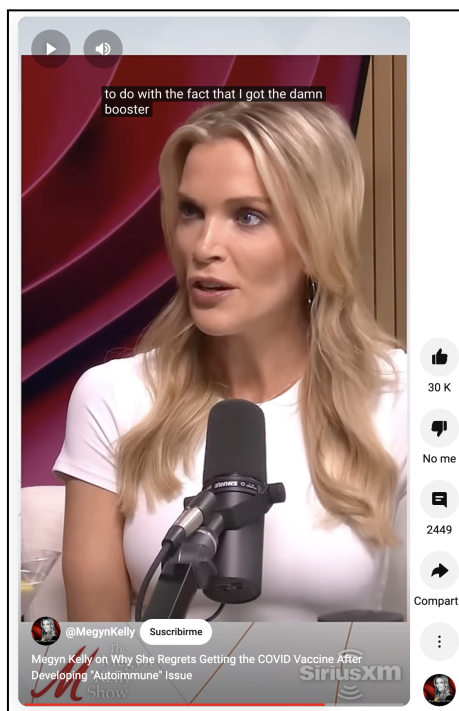


In YouTube, the post with more visualizations has **over half a million reproductions** and claims that “[EU Commission, President Ursula von der Leyen, has lifted the recently decided extensive ban on new cars with internal combustion engines from 2035](#)”. The video is from a German youtuber and gathered **over 5,000 comments**. The post not only does not show any visible action, but also is accompanied by **ad banners and in-feed video ads**.

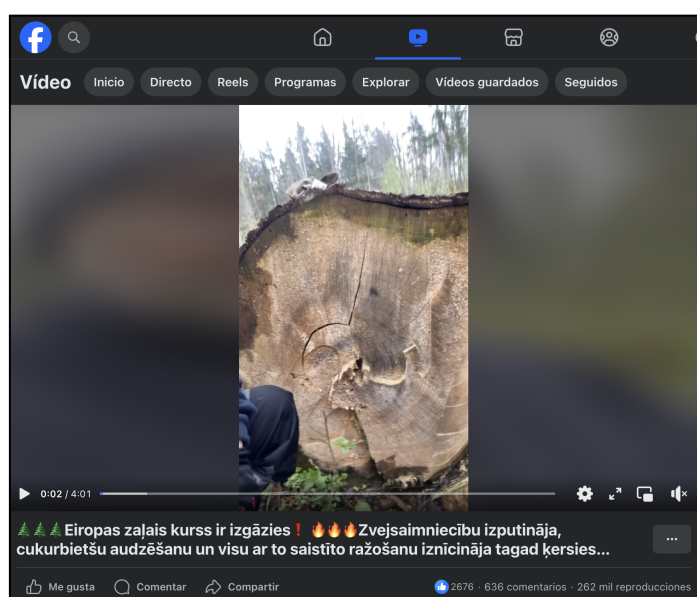


Also in YouTube, one of the most viral videos that did not receive any visible action had **over 300,000 reproductions** claiming that “[COVID-19 vaccination is associated with a higher risk](#)”

[of autoimmune rheumatic diseases](#)". The video, in a reel format published in Poland, summed up almost 2,500 comments.



On Facebook, one of the most viral debunked posts that did not receive any visible action by the platform is a video published in Latvia with **over 250,000 reproductions**, claims that "[If CO2 in the atmosphere fell below 0.03%, then all living things would die. So the green course is a complete miss](#)". The video has over 600 comments.



Another of the most viral posts is a reel video with over 12,000 likes published in Austria claiming that “[The US Senate has announced that Corona is a lie and that pharmaceutical companies are lying](#)”. The video has more than 300 comments.



In Instagram, debunked posts with no visible action have, in general, less virality than in the other platforms. One of the most viral debunked posts is a video published in France with over 16,300 reproductions and almost 1,300 likes claiming that “[France will send citizens to jail if they question mRNA vaccines](#)”.



Another viral video debunked post in Instagram is a video published in Spain that claims that [“Bulgaria has ceased to be a member of the World Health Organization”](#) and accumulates over 68,000 preproductions and almost 3,900 likes.



5. Methodology

In order to evaluate how platforms are responding to disinformation content that might affect the European Parliamentary elections, we have analyzed the reported “false”, “partially false” and “AI generated” content identified by the fact-checking organisations involved in the [Elections24Check](#) project during the four months ahead of the elections, including the Election Day. Elections24Check is a database that gathers and categorizes disinformation and connects it to the evidence provided by fact-checkers on why a particular content is deemed disinformation.

For this analysis, we manually checked one by one the platform responses to each piece of disinformation content collected in the repository. The platforms selected for this report have been Facebook, Instagram, TikTok, X and YouTube. In total, we identified 1321 pieces of content across those services.

Platform responses have been classified as follows:

- **Removal:** When the content does not appear anymore, either because it was removed by the user, or because it was removed by the platform.
- **Suspended account:** when the content does not appear anymore following a suspension of the account by the platform. This notification has only been identified in X.
- **Debunk label:** When the label includes a link to a debunk made by independent fact-checkers. This kind of label has only been identified in Facebook and Instagram.
- **Generic label:** When the label does not include a link to a debunk made by independent fact-checkers. It usually includes generic information such as noting that multimedia content is presented out of context and linking it to the terms of service of the platform. It has been only identified in X.
- **Community Note:** X’s community-driven content moderation notes.
- **Label + Note:** When a content includes both a generic label and a Community Note. The generic labels and the Community Notes counted here are not counted separately.
- **Information panel:** Notes linking reliable sources of information based on the topic of the content published. It has been identified in TikTok, linking to “info on the European elections” from a dedicated website created by TikTok; and in YouTube, linking to information about climate change from the United Nations.
- **Limited visibility:** When content visibility has been limited and it is notified next to the post. This notification has only been identified in X.
- **No action:** When no visible action is shown in the post. This may include non-apparently visible actions, such as reducing content visibility, where this is not flagged to the public in the post.