



Digital Platforms' Responses to Disinformation in the Spanish 2023 General Election

An analysis of the actions taken by six monitored VLOPs on disinformation identified by Maldita.es during the Spanish general election campaign of July 23, 2023



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

Summary	3
Service by Service Analysis	5
YouTube	5
Facebook	6
Instagram	7
TikTok	8
X (Twitter)	9
LinkedIn	10
Comparison of responses	11
Methodology	13

Summary



This report analyses the response of six very large online platforms to the most viral disinformation contents identified by Maldita.es on their services during the campaign for the Spanish general election of July 2023. Among the 57 top-performing disinformation posts previously debunked by Maldita.es, a concerning 80.70% (46 posts) elicited no visible action by the platforms. This highlights the need for these services to improve the way they address disinformation during electoral processes.

Platform by platform, we see that among the top-performing posts, Facebook stood out, with 8 out of 10 receiving effective actions, primarily in the form of fact-checking labels linked to evidence. However, on Instagram and LinkedIn, no content seemed to have undergone any action. For TikTok and YouTube, only one post was removed, and no apparent action was taken on the remaining content. Similarly, on platform X, only one post showed a Community Note, while the rest showed no signs of action taken.

28% of all top-performing disinformation posts across platforms featured attacks on the integrity and fairness of the electoral process, while 20% included falsehoods about climate change. All major parties had posts targeting them among those most viral during the electoral period. Those 57 most viral posts gathered more than 250,000 interactions and were seen by millions of users.

In the first weekend of the official campaign, a post on X falsely said that the EU Commission president [Ursula Von der Leyen was on the campaign trail supporting one of the candidates](#), a post that eventually reached 1.2 million people and was also among the top-10 disinformation contents on YouTube and Facebook. Two days before the election, a government minister and candidate falsely [accused a rival party of wanting to kill all stray dogs if they were elected](#), a post seen by 1.9 million.

Almost 500,000 people watched on TikTok [an old video that was resurfaced to promote unsustainable accusations of fraud regarding vote by mail](#), and 38,000 saw one on YouTube that [falsely accused postal workers of throwing into the trash the votes of a particular party](#). Millions saw an allegation against the climate minister [falsely accusing her of taking a government jet to a close-by climate summit](#), a disinformation that found its way into all the very large online platforms monitored, reaching more than 130,000 on Instagram alone.

At the time of the election, all those services were already designated as very large digital platforms (VLOPs) under the EU's Digital Services Act, but not yet subject to its full obligations including that of addressing any *"actual or foreseeable negative effects on democratic processes, civic discourse and electoral processes"* by deploying effective risk mitigation measures. The Spanish election was therefore the last big test before the law fully entered into force for VLOPs in the European Union.

Service by Service Analysis



YouTube

Of the top-performing disinformation posts identified by Maldita.es on Youtube, only 1 of the 9 posts with further engagement might have been treated by the platform, given that YouTube favours removals as opposed to warnings or contextual interventions for disinformation. By 10 November 2023, only one post appeared removed while the rest were still visible and showed no labels more than three months after being debunked by Maldita.es. The most viral post has over 50,000 visualisations and was shared by an account with almost 25,000 subscribers.

Several videos are still unlabelled, and the most viral one is also preceded by advertising and with in-feed ads (an ad that appears in the 'video discovery' section of Youtube, copying the format of the videos recommended by the platform) next to it. This is an example of disinformation content benefiting from poor control of monetisation and ad placement, and it shows the need for Youtube to improve their policies and systems in this area.



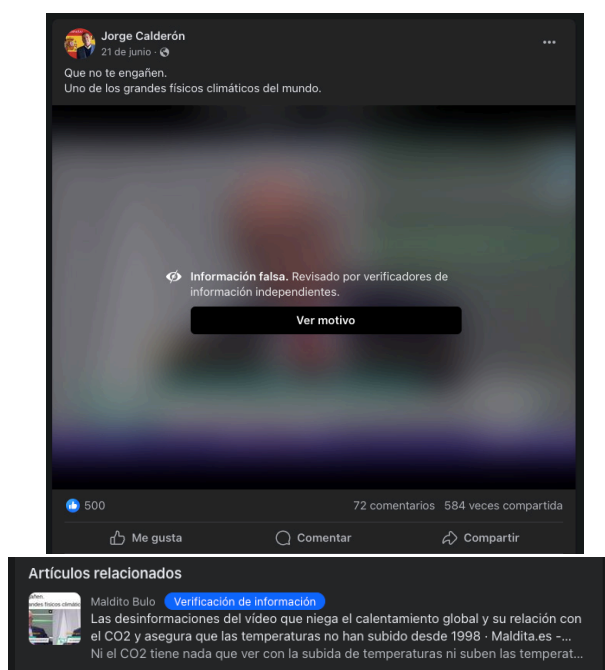
Key takeaways:

- YouTube has only taken action in 1 out of 9 posts.
- The only measure taken, if any, is the removal of content.
- The most viral post is still being monetised, containing ads both in-stream and in-feed.

Facebook

Of the top-performing disinformation posts identified by Maldita.es on Facebook, 8 out of 10 received some kind of action. 6 of the posts appeared labelled as “false” or “partially false information” and linked to the evidence on why an independent fact-checker had rated them as such, while 2 posts appeared removed (perhaps by their authors after being notified of a fact-checking label). The other 2 were still visible in November 2023.

This is a positive example of the effectiveness of disinformation labelling in online platforms, as out of a larger sample of 489 Facebook posts containing electoral disinformation debunked by Maldita, just 7% had more than 5 reshares and less than 1.5% had more than 25.



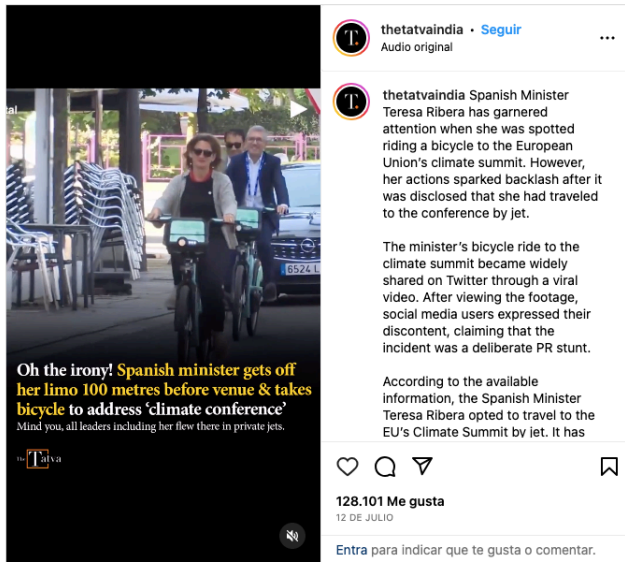
Key takeaways:

- 8 out of the 10 top-performing posts received some kind of action, including labelling.
- Facebook is the service that has taken more effective measures against the disinformation content identified by Maldita.es in this analysis.
- Facebook’s third-party fact-checking programme allows partner organisations to proactively flag contents as containing disinformation (including Maldita.es), resulting in the labelling of that content so users can access the evidence provided by the fact-checkers.

Instagram

Partly because of the current difficulties to monitor disinformation claims through searches on mainly video-and-image platforms, we identified only 8 Instagram posts with significant engagement that contained electoral disinformation. By 10 November 2023, more than three months after being debunked by Maldita.es, the 8 posts still appeared visible and unlabelled, with 3 of them having accumulated more than 1,000 likes each and being posted by accounts ranging from 40,000 to 290,000 followers each.

Despite Facebook and Instagram both belonging to Meta and thus sharing similar features to integrate fact-checking into their services, the difference between the results on Facebook and Instagram is more than evident. Improvements in the monitoring tools are much needed to boost the fact-checkers capacity to proactively spot and flag disinformation, as well as improvements in the automated matching that Meta undertakes to scale the program.

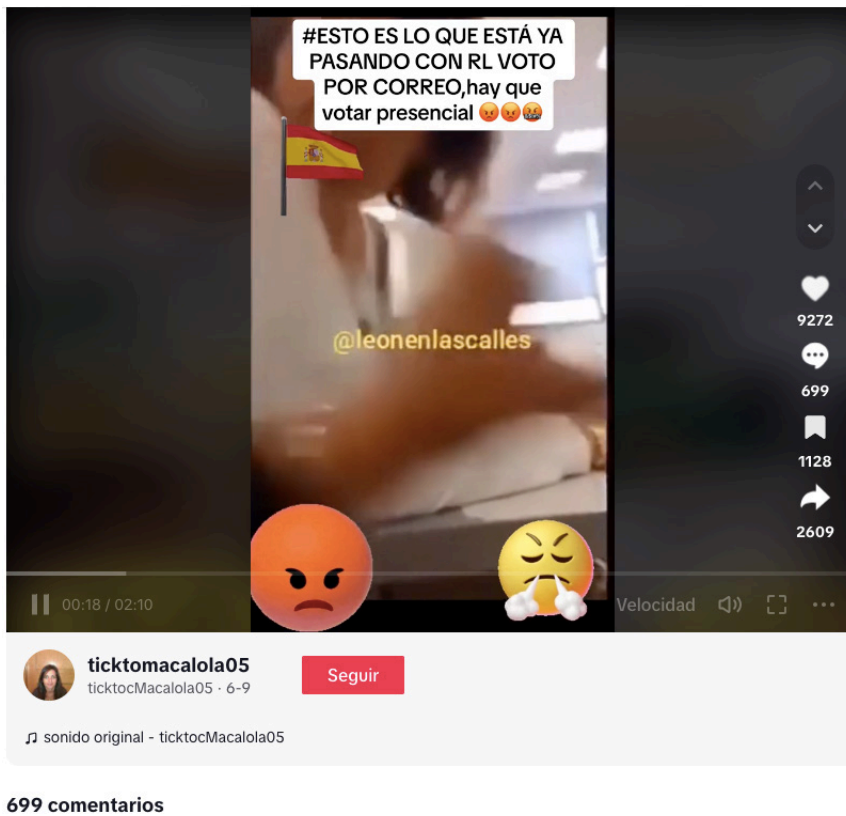


Key takeaways:

- Instagram has taken no action on 8 out of 8 posts.
- Instagram is the service that has taken less effective measures against the disinformation content identified by Maldita.es.
- Although the integration of fact-checking on Instagram is designed to be similar to that on Facebook, the results show a very different implementation on both platforms.

TikTok

Of the top-performing disinformation posts identified by Maldita.es on TikTok, only 1 in 10 may have received some kind of action, because by 10 November 2023, 9 of them still appeared visible and unlabelled, despite one of them reaching over 67,000 visualisations and 23,700 reshares, and five of them accumulating more than 1,000 interactions each.



Key takeaways:

- TikTok has not taken action in 9 out of 10 disinformation posts, at best.
- No “unverified” labels were applied to any of those contents.

X (Twitter)

Of the top-performing disinformation posts identified by Maldita.es on X, only 1 of the 10 posts was affected by moderation measures from the platform. By November 2023, only one post included a Community Note giving context and debunking its content. However, the rest still appeared visible and unlabelled. This is the case more than three months after being debunked, and despite 70% of them reaching over 1 million visualisations and 20% over 11.5 million by the time they were identified by Maldita.es.

This proves that X's Community Notes model is limited in its capacity to address the spread of disinformation, even for highly viral posts and in environments where the platform has a big user base, as Spain is the biggest EU market for X with 22 million monthly users. This might also be connected to the low number of Spanish speakers working in content moderation for X (just 20 out of 2,294), reflected on X's DSA transparency report.



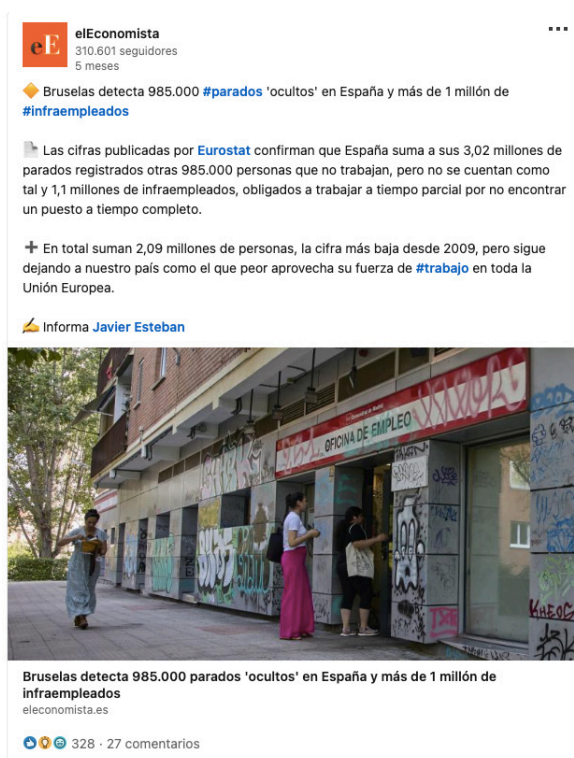
Key takeaways:

- X has only taken action in 1 out of 10 disinformation posts.
- The only measure taken is one Community Note, despite the huge virality of the rest of the content.

LinkedIn

LinkedIn was not part of our original electoral monitoring, but the viral disinformation debunked by Maldita.es during the electoral period was still present on the platform several months after the election when we looked for it to produce this report. There is no way of knowing whether other similar contents did receive action from the platform, but it seems unlikely as LinkedIn reports that [it only fact-checks around 1,500 contents annually](#) in all the world.

We have identified 10 posts around the period of the elections containing different content catalogued by Maldita.es as disinformation with no apparent action by the platform. At least some of them were quite viral, with one reaching 328 reactions, 27 comments, and 52 reposts.



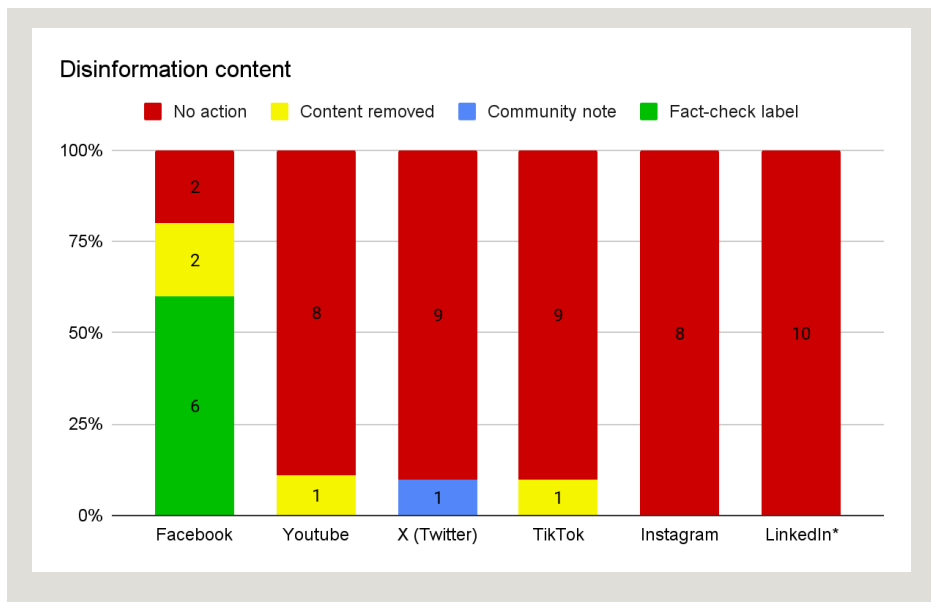
Key takeaways:

- 10 posts containing disinformation are still available after more than 3 months, being 2 of them relatively viral.
- None of these posts display any type of fact-checking.

Comparison of responses

When comparing the responses and measures taken by the six services, we see that Facebook has been the most effective, addressing 80% of the top-performing disinformation content identified by Maldita.es. Video-based platforms, namely TikTok and YouTube, have a similar removal rate, with 10% and 11% of content removed, respectively, with no labels or warnings added to the remaining posts. X also indicates little success with 10% of posts featuring Community Notes, while the rest saw no action. Instagram lags behind in addressing the content debunked by Maldita.es, with no apparent measures taken in any of the posts.

Regarding LinkedIn, despite the fact that there is no way of knowing whether similar content was removed before this report was done, the current evidence shows that none of the disinformation contents identified by Maldita.es incorporates any form of fact-checking for users, even three months after being debunked.



Although the positive results of Facebook might be explained by the long-term structured collaboration with fact-checkers, any other visible measures (for example, X's Community Notes) were also considered for the study and found little efficacy.

Moreover, such cooperation with fact-checkers is a commitment subscribed by all the platforms contemplated in this study as part of their participation in the Code of Practice on Disinformation, although X reversed course and left the Code in March 2023.

Despite the results of this analysis being based on a limited number of disinformation contents, they show how relatively (and in some cases, hugely) viral contents have not been effectively addressed by most of the services even months after being published and debunked by Maldita.es.

This serves as evidence that during the electoral process in Spain, a short few months before the DSA came into force for them, most of the very large online platforms had not taken sufficient "effective mitigation measures" against disinformation that presents "actual or foreseeable negative effects on civic discourse and electoral processes".

Methodology



For this study, we selected for each platform the 10 posts that experienced further engagement among those that contained electoral disinformation already debunked by Maldita.es, from a total of 781 posts, and compared the different actions taken by each platform in regards to them, including when no action was taken. The analysed digital platforms are six of the “Very Large Online Platforms” (VLOPs), designated as such according to the European Union’s Digital Services Act (DSA): Facebook, Instagram, LinkedIn, TikTok, X (previously Twitter), and Youtube.

The original database used for this study included posts collected from 24 June to 23 July 2023, all of which included claims that had already been debunked by Maldita.es. To collect these posts, we combined manual searches directly conducted on the five platforms initially analysed and links reported through Maldita’s WhatsApp chatbot and other tiplines. Three months later, a second round of searches was conducted to find out whether the platforms had taken any action on the identified disinformation contents. During that second round, we also checked if well-known election-related disinformation claims published before the election were still present on LinkedIn three months after the vote, as that platform had not been monitored during the first round.

Where 10 posts with high engagement were not identified on a single platform, only the most viral were considered. All of the identified disinformation had already been debunked, so in 100% of the cases platforms could have provided additional responses to the disinformation present in their services.

If you have any questions, please reach out:

policy@maldita.es