
YouTube and TikTok' Measures Against Climate Disinformation: How to Repair an Obscure System that Doesn't Work

Disinformation about climate change can have extremely harmful consequences: it negatively affects preparedness for natural disasters and can unnecessarily undermine public confidence in evacuation orders and other public service announcements during extreme weather events. It can also prevent the adoption of effective public policies to address climate change, when presenting it as innocuous or unrelated to human activity, if not outright denying its existence.

Those dangers belong in some of the systemic risk categories outlined in the EU Digital Services Act, namely those with “actual or foreseeable” negative effects on “civic discourse and electoral processes”, “public security”, “the protection of public health”, or “the person’s physical and mental well-being”. The same regulation obligates platforms over a certain size (including YouTube and TikTok) to have effective measures in place to mitigate those risks.

Most major digital platforms recognize those risks to an extent and have provisions in their terms and conditions to limit the damage. [In a study](#) we just published alongside Science Feedback, we analyzed how YouTube and TikTok apply their climate misinformation policies in Spanish and French, and how out-of-court dispute-settlement bodies (ODS) or national digital regulators responded to complaints about the platforms’ moderation decisions when those do not align with their own policies. The results are striking.

YouTube, whose advertising rules prohibit the monetization of content that contains climate dis/misinformation, seldom applies that policy even when alerted that a given monetized video is in breach. TikTok, which in theory does not allow climate dis/misinformation that contradicts scientific consensus, applies it very inconsistently. When bringing the complaints to the ODS Appeals Centre Europe and the French DSC Arcom we found the response times were long and could document the limitations of the process. But along the way, we learned a few valuable lessons.

Moderation for Climate Dis/misinformation is poorly designed, seldom applied, and unfit for purpose

The larger problem with **TikTok** goes beyond its inability to apply its policies, it affects its rules themselves. Takedown is not an appropriate action for disinformation unless it poses an immediate risk to people; a better intervention is to label, demote and/or demonetise content. Labeling alerts the public of false information without infringing on freedom of expression whereas demotion and demonisation reduce the spread of disinformation or disincentivizes its creation. Instead, removal can be perceived as censorship.

Our results show those considerations might also be affecting TikTok's implementation of its policies: even on the 20% of reported videos containing climate dis/misinformation that TikTok did act upon, the platform chose to demote them instead of removing them. An analysis of the content that was moderated indicates that TikTok was applying a different rule, one against conspiracy theories, rather than its main policy on climate dis/misinformation that many violated and that triggers content removal. As much as we consider that policy a mistake, we believe platforms have an obligation to enforce their policies faithfully as required by article 14 of the DSA.

On **YouTube**, we consider that its policy of not remunerating the creation of content that contains climate dis/misinformation is positive, even if insufficient. YouTube should definitely label content to give users additional information when seeing false claims, but even if such a change would be adopted, the fact that our analysis found that YouTube does not apply its current policies on the matter at all makes us think that whatever the policies, lack of enforcement makes them irrelevant. After all, even when alerted by us, YouTube barely acted on 4% of the violative content reported.

The lack of transparency about enforcement creates problems beyond transparency itself

Our position is that platform response to disinformation needs to be focused on offering additional context rather than content removals, but also that any moderation action on that area must prioritize transparency and verifiability of enforcement for content that is legal but harmful. That said, when penalties below deletion are applied and removal of content is mostly out of the question (as it should be) enforcement relies on measures like demotion and demonetisation and those actions are challenging to monitor. Whereas a takedown video leaves a trace, a demoted video does not. Because of this, there is little information accessible to determine whether platforms' own policies are truly followed.

This makes holding platforms accountable harder and complicates research on DSA systemic risk, particularly in an era when Big Tech have [aggressively stonewalled](#) researcher data access. When it comes to the relevant YouTube and TikTok climate dis/misinformation policies establishing penalties below removal, they do not provide enough data to assess whether or not their policies are being enforced. In these cases, researchers have to fall back on proxies like we did. For example: imagine clicking through over 80 videos to check whether, sometime after reporting them for violating YouTube's policies, they still show ads before playback, which would indicate they continue to be monetised.

Some penalties across platforms are even harder to account for since there are no obvious proxies: there might be videos whose visibility have been reduced as the result of a moderation decision based on climate dis/misinformation policies but even the user who reported them in the first place might have no way to know if that is the case. DSA's Article 20, that says complainants of a platform moderation decision (for example the decision to demote a video or not) are entitled to a response to their report, but YouTube for example offers none for decisions that do not involve a content removal.

While TikTok and YouTube's DSA transparency reports offer aggregate numbers, those are not helpful for an actual independent check on their enforcement. [Tiktok](#) says 64% of the contents they moderate under their misinformation policy were removed and 36% had their visibility restricted, while [YouTube](#) listed more than 153,000 moderation actions under its ads policy, all of them not involving a visible action. These cannot speak to the accuracy of the enforcement action or its effects on users' exposure to this type of disinformation.

ODSs need more resources and “good faith” from the platforms

Only 10% of the complaints against moderation decisions (undermoderation, in this case) that we directed to the ODS Appeals Centre Europe were decided within a month. This waiting period is inadequate to effectively address the harms created by the content even if platforms were willing to reverse course as a result of the ODS decision. The blame for this delay is not necessarily on the ODS, as they have often described the limitations they encounter from platforms in accessing the information they need to decide on complaints.

The regulation does not make the ODS decisions binding but requires platforms to cooperate with them “in good faith” and lists the implementation of the ODSs' decisions as a way of mitigating systemic risks. Now, if the ODSs themselves are kept in the dark in respect to the real impact of their rulings, that not only increases the opaqueness around content moderation decisions below the takedown threshold (and taken as a result of users' reporting) but it also complicates any independent assessment of that particular risk mitigation tool.

Platforms need to produce **data on individual complaints** and regulators must continuously assess whether the platforms' engagement with the bodies is consistent with the principle of 'good faith' established in the regulation. Although not binding, platforms should be expected to explain why they do not implement ODS decisions and if the data indicates they do so at scale, that should feed into their compliance assessments. As [Husovec aptly puts it](#), "the effect of ODS decisions is best described as 'systemically but not individually binding". This implies that while platforms can dismiss specific requests, they shouldn't be able to consistently overlook the mechanism as whole without justification.

Higher transparency should also be required when it comes to monetisation, making available across all platforms which accounts participate in revenue-sharing programs, and how revenue is shared across creators, creating specific pathways for demonetisation in internal complaint-handling mechanisms and informing users when a report results in demonetisation of the violative content.

[Access to the study by Fundación Maldita.es and Science Feedback.](#)