META: STILL NOT DSA-READY

conspiracy theories

HATE SPEECH

election disinformation

violent uprisings

EKŌ

As DSA goes into effect, Meta approves series of violent, racist, anti-semitic, and stop the steal ad content targeting Europeans

ONE OF THE ADS CALLED FOR THE EXECUTION OF A PROMINENT MEP FOR THEIR STANCE ON IMMIGRATION. HERE'S HOW THE DSA COULD CRACK DOWN ON META'S HARMFUL CONTENT.

New research by corporate accountability group, Ekō, in collaboration with the People Vs. Big Tech network shows that Meta is still failing to detect and block ads containing hate speech, election disinformation and incitement to violence – including a death threat against a sitting member of the European Parliament.

The alarming findings come the day before the EU's flagship Digital Services Act (DSA) comes into effect, which if enforced properly will target the core of Big Tech's broken business model that supercharges the spread of hate speech and disinformation.

In an experiment carried out between 4th-8th August, Facebook approved a series of eight highly inflammatory ads calling for a stop the steal-style violent uprising to overturn the results of the recent Spanish election, racist and anti-semitic slurs, and incitement to violence against immigrants and the LGBTQ+ community.

Each ad text was accompanied by manipulated images generated by AI image tools, showing how quickly and easily this new technology can be deployed to amplify harmful content.

META ADS ARE MONETIZING CONTENT CALLING FOR EXECUTIONS, GENOCIDE AND STOP THE STEAL

In total 8 out of 13 ads were approved by Meta within 24 hours; all of the approved ads broke Meta's own policies. Five ads were rejected on the basis that they referenced elections or politicians and therefore were political ads. All of the ads were removed by the researchers before publication, meaning they were never seen by Facebook users.

- Several ads played to fears of Europe being swamped by immigrants and linked immigration to alleged instances of violent crime.
- An ad geo-targeted at an audience in Germany called for synagogues to be burnt to the ground to 'protect White Germans.'
- Two ads pushed a 'stop the steal' narrative around the recent election in Spain, claiming electronic voting machines were rigged and called for a violent uprising to murder political opponents and reverse the election outcome.
- One ad geo-targeted in Romania called for the cleansing of all LGBTQ+ identifying people.
- One ad called for the execution of a prominent MEP because of their stance on immigration.
- Each ad was accompanied by a manipulated image created with the AI image tools, Stable Diffusion and Dall-e2. For example, Ekō researchers were able to easily generate images showing a masked person stuffing ballots into a ballot box, drone footage of immigrants crowding at ports and border crossings, and synagogues on fire.

The ads were placed in German, French, English and Spanish. The researchers removed the ads before publication meaning they were never seen by Facebook users.

An additional five ads were submitted and rejected on the basis that they may qualify as political ads, but they were not rejected on the basis of hate speech or inciting violence, which they contained. Meta requires accounts running political ads to go through a specific authorization process to verify the account holder's identity, as well as applying some further restrictions on the political content of ads in specific regions at specific times, for example during elections.

Alarmed by the global backsliding of democracies, and the growing success of far right and anti-democratic actors in disrupting elections in the United States, Brazil and Kenya in recent years, civil society groups, led by the People Vs. Big Tech network, are increasingly worried about the threat disinformation and hate speech poses to the upcoming 2024 European elections. Facebook, YouTube, TikTok and other social media sites have enabled groups to easily seed and amplify election disinformation and conspiracy theories, sometimes ending in real world violence and even attempted coups. Over 50 civil society groups are urging the European Commission to take pre-emptive action and use its powers under the new Digital Services Act to force companies to account for how they will stop the flood of election disinformation.

WHEN WILL META GET A GRIP ON HATE SPEECH AND DISINFORMATION ON ITS PLATFORM?

The ads highlight once again Meta's toxic business model and sub-standard moderation practices which have upended elections and fuelled real world political violence. Each ad that was approved in this experiment was in clear breach of Meta's own policies, and demonstrates that the systems Meta has in place to detect extremist and violent content are not fit for purpose.

This investigation follows a string of reports exposing Meta's failures to protect users in regions across the world. Recent Ekō research in Brazil uncovered an ecosystem of ads and posts peddling conspiracy theories about the integrity of the election and supporting far-right calls for a coup. Global Witness investigations have also shown how Meta is failing to detect ads containing hate speech and electoral disinformation in Myanmar, Kenya, Ethiopia, Brazil and the United States.

Despite the wealth of evidence of systemic failures and real-world harms, Meta has failed to take substantive corrective measures. Ads containing highly inflammatory hate speech, violent intent and disinformation are still being greenlit by its ads approval system.

HOW EUROPE'S DIGITAL SERVICES ACT WILL CURB HATE SPEECH AND DISINFORMATION

From August 25th, the world's biggest platforms, known as Very Large Online Platforms (VLOPs) will be legally required to comply with the Digital Services Act (DSA). EU Commissioner for the internal market, Thierry Breton, visited Silicon Valley earlier this summer and met with Meta who assured him that the company was ready to meet its obligations under the DSA. But this latest research indicates Meta is still falling well short of what is needed to comply with these new laws; and with AI generation tools being developed and rolled out for commercial use at lightning speed, there is potential for disinformation to spread at an unprecedented scale and speed.

Under the new law, tech companies will have to make their platforms safe-by-design by assessing and mitigating against systemic risks in the design and roll-out of their products and services.

The DSA defines 'systemic' by referring to 'actual or foreseeable negative effects' on the exercise of fundamental rights, dissemination of illegal content, civic discourse and electoral processes, public security and gender-based violence, as well as on the protection of public health and minors and physical and mental well-being. This includes systemic risks posed by coordinated disinformation campaigns by state-sponsored actors or extremist groups, or by platform users pushing climate disinformation. The series of ads approved in this experiment clearly show that Meta is falling short of identifying and mitigating these types of systemic risks.

SPECIFICALLY, HOW THE DSA COULD ADDRESS THE SCENARIOS DOCUMENTED IN THIS REPORT

If the ads in this research went live, the DSA would provide several different measures to address it. First, as mentioned above, the platform is required to identify and mitigate systemic risks. In this case, Meta's frictionless content moderation system which approved a series of disinformation and illegal content, is a systemic risk that the platform will be required to fix under the DSA obligations. But without DSA enforcement, the platform has very little incentive to address these systemic risks, especially as its business model depends on advertising and amplifying all types of content – including the kind that drives high engagement like hate speech and disinformation.

The platform would also be forced to report on how it uses automated content moderation tools as well as a tool's error rates.

This would expose the platform's failure in a continuous manner instead of relying on groups like Ekō and other civil society to expose its failure.

The DSA would also force Meta to disclose all the notices it receives of illegal content whether it's from trusted flaggers or from automated systems. This would expose the sheer scale of illegal content circulating on the platform.

This push for transparency will become the basis for holding the platforms accountable for spreading disinformation and illegal content. This could lead to fines levied in the billions. But accountability will only happen if EU authorities actually enforce the DSA.

HOPE FOR DEMOCRACY: HOW THE DSA COULD SAFEGUARD UPCOMING ELECTIONS IN THE EU AND ELSEWHERE

Social media platforms, with their toxic algorithms are fuelling the massive growth of online hate speech and disinformation, and extremists are using these sites to facilitate and amplify disinformation and conspiracy theories to challenge election outcomes. 2024 is a pivotal year for elections with over 50 countries going to the polls including India, US and Europe. The European Commission has a critical opportunity to force platforms to bring their operations in line with democracy and human rights.

Civil society groups are calling on the European Commission to use its powers under the DSA to require companies to publish detailed plans of how they will deal with disinformation and other risks during these upcoming national and EU elections. Their list of demands from the platforms include:

1. Deamplify disinformation and hate speech
Tech platforms have shown they can switch on measures to make content less viral at critical moments. They must, as a matter of course:

- Make their recommender systems safe-by-design, by default and all the time (not just during election periods), including measures to suppress the algorithmic reach and visibility of disinformation and hate-spreading content, groups and accounts.
- Implement meaningful user control features, including giving users clear options to choose over which types of data are used for ranking and recommending content and the ability to optimise their feeds for values other than engagement.

2. Ensure effective content moderation in every European language

The tragic impacts of viral hate speech in Ethiopia, Myanmar and countless other places shows content moderation is worthless if not properly and equitably resourced. Tech platforms must:

- Properly resource moderation teams in all languages, including both cultural and linguistic competency.
- Make content moderation rules public, and apply them consistently and transparently.
- Pay moderators a decent wage, and provide them with psychological support.

3. Stop microtargeting users

The potential to exploit and manipulate voters with finely targeted election disinformation is an existential danger for democracy. The solution is to:

- End processing of all observed and inferred data for political ads, for both targeting and amplification. Targeting on the basis of contextual data would still be permitted.
- Enforce the ban on using sensitive categories of personal data, including data voluntarily provided by the user, for both targeting and amplification.

4. Build in transparency

Elections belong to us, not social media companies. Tech platforms must not be allowed to shape the fate of elections behind closed doors – instead, they must:

- Be fully transparent about all measures related to political content and advertisements, including explanations of national variations in the measures they put in place, technical documentation about the algorithms used to recommend content, publication of ad libraries and their functionality (as well as ad financing) and full disclosure of content moderation policies and enforcement including notice, review and appeal mechanisms.
- Allow researchers and wider civil society to independently monitor the spread of dis/misinformation and potential manipulation of the information space by sharing real-time, cross-platform data, including: content meta-data; information on content that is demoted, promoted and recommended, and tools to analyse data.
- Provide training for researchers, civil society, independent media and election monitors to monitor activity on the platforms.
- Facilitate independent audits on the effectiveness of mitigation measures adopted in the context of elections and publish their results.

5. Increase and strengthen partnerships

Companies are not experts in elections. They must work with those who are.

- Companies must meaningfully engage with partners such as fact-checkers, independent media, civil society and other bodies that protect electoral integrity, taking into account partners' independence and reporting on their engagement in a standardised format.