AVAAZ

Left Behind: How Facebook is neglecting Europe's infodemic

As Europe enters a third COVID-19 wave, our research finds that Facebook is failing to protect Europeans from dangerous misinformation. Facebook promised to do more to protect its users, but a year after the start of the pandemic, our findings suggest it has not improved its ability to detect dangerous misinformation, emphasising the need for urgent EU regulations.

April 20, 2021

Contents

Executive Summary

Facebook's "America First" approach to fighting misinformation fails to protect European citizens

- A majority (56%) of fact-checked misinformation content in major non-English European languages [1] is not acted upon by Facebook, compared to only 26% of English-language content debunked by US-based fact checkers. [2]
- This means Europeans are at greater risk of seeing and interacting with COVID-19-related misinformation:
  - Italian speakers are least protected from misinformation, with measures lacking for 69% of Italian content examined. [3] Next are French and Portuguese speakers, with measures lacking on 58% of French content [4] and 50% of Portuguese content. [5] Spanish speakers were most protected, though measures lacked for 33% of Spanish language content, [6] which is still more than English content.
- **Based on our sample analysis, on average, Facebook is almost one week slower to label non-English false content,** taking 30 days for that content compared to 24 days for English-language false content.

Facebook's fight against the infodemic has failed to make an impact one year on

- Despite a number of commendable steps to fight the ongoing infodemic, **Facebook's action on fact-checked misinformation slightly decreased** compared to last year [7] (55% in 2021 vs 56% in 2020).
- **The average delay between post publication and Facebook labelling was** 28 days in 2021.
- The platform fails to detect and label exact copies, **"clones"** or slightly altered, **"variants" of false claims**, i.e. minor changes to format or narrative on claims that have been debunked and already labelled on the platform.

- **51 such "clones" and "variants" reached an estimated 807,746 total interactions, and 63%** of this content is missing warning labels by Facebook. [8]

The top misinformation narratives we identified risk boosting vaccine hesitancy, and deter mask use

- The top misinformation narrative we identified (36 posts, 1.4M interactions) is about vaccine side effects, including death. One example is this false item about Bill Gates , who was claimed to have said during an interview that "[...] for every 10,000 people there would be permanent vaccination damage, including 700,000 expected deaths". [9] **(29,050 interactions)**
- The second top COVID-19 misinformation narrative was around false official measures or warnings (16 posts, 536K interactions) See example here .
- The third top narrative was against masks, claiming they are either dangerous or useless. As an example, harmful misinformation claims that using masks leads to cancer and other diseases. We found 12 examples of this "clone" or "variant"- 92% unlabelled. See two examples here and here . Together, the 12 posts [10] have reached **118,288 interactions.**

EU must roll out strong regulation to inoculate against the infodemic

- Social media self-regulation has failed: one year into the ongoing pandemic, **Facebook has not kept its promise to fully identify COVID-19- and vaccine-related misinformation in Europe.**
- The current EU Code of Practice on Disinformation does not cover the failures identified in this report. That is why we urgently need a revised version that pushes social media giants to disclose the amount of misinformation on their platforms and set clear goals for its reduction, monitored by an independent regulator.
- A new code will only be effective if it asks social media platforms to notify all users who have interacted with misinformation that they have done so and to reduce the algorithmic acceleration of misinformation content and actors.
- The analysis in this report is based on a sample of misinformation detected by our investigation team that was also fact checked by independent fact checkers. Facebook is not transparent and does not provide data, which would allow for a more detailed analysis on the full scale of misinformation on its platform. The EU Code of Practice must require more transparency from the platform to help researchers understand the full reach and harm caused by such content.

Key Findings

For this study we have analysed 135 pieces of misinformation content in five different languages (English, French, Italian, Spanish and Portuguese) that were posted in late 2020 or early 2021 and were rated false and misleading by reputable, independent fact checkers. In the graphs that follow we compare our analysis with the one we performed in 2020 , using exactly the same methodology used in the report we released in the pandemic's first months. For more information please refer to the Methodology and Data section.

Figure 1 shows how many of the posts analysed were either labelled, removed or remained unactioned on the platform during the period of our investigation. From the image, you can see how there is no improvement from last year, with actually a slight decrease in posts detected by Facebook. Another main difference is that twice as many posts were removed, which would be a positive indication that Facebook was acting more robustly, however we have no way to know if the removal was done by Facebook or by the account which had posted the content. It is important to note that Facebook announced on February 8, 2021 , after years of pressure from health experts and civil society organizations, that it would ban misinformation about all vaccines. Nonetheless, Avaaz continued to find such content on the platform.

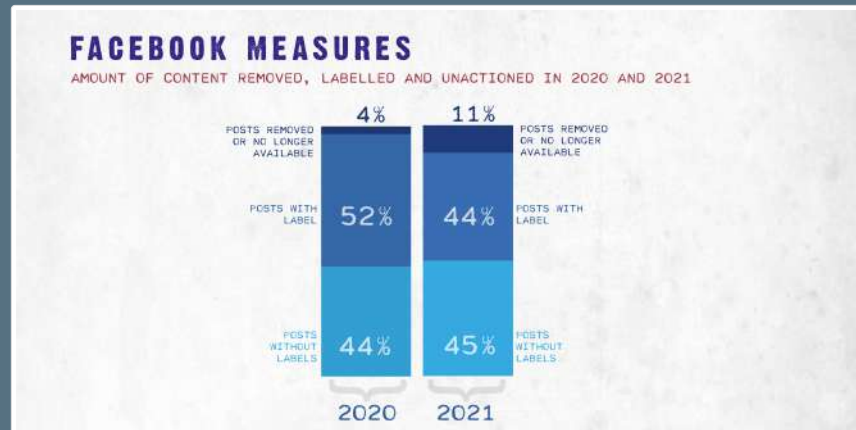Moreover, there is also a significant reduction in posts that remain on the platform with a label.

Figure 1 - Comparison between the amount of content removed, labelled and unactioned in our 2020 and in our 2021 analysis

In Figure 2 we break down the same data per the different languages analysed in this report. Italian and French seem to be the most neglected languages. On the opposite side, Spanish is the non-English language with the highest amount of actioned content, potentially as a consequence of the media focus it had during the recent US elections and of the work of civil society groups, including Avaaz, in flagging that Spanish language misinformation was a problem [11] .

Figure 2 - Comparison between the amount of fact-checked misinformation content unactioned in our 2020 and 2021 analysis. Breakdown per language

Figure 3 shows the comparison between English and non-English content. It can be seen that a majority (56%) of the content remains unactioned in major non-English European languages [12] , compared to only 26% of English-language content debunked by US-based fact checkers [13] . This leaves all users in Europe at a greater risk of seeing and interacting with COVID-19-related misinformation without any fact-checking measures.

CONTENT WITHOUT MEASURES

ENGLISH VS. NON-ENGLISH LANGUAGES

ENGLISH (US) — 26%

ENGLISH (UK & IRELAND) — 50%
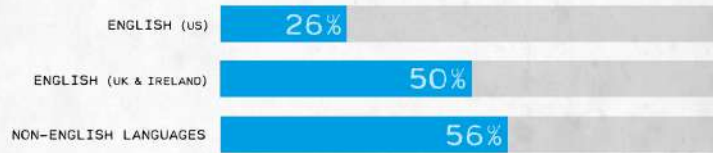
NON-ENGLISH LANGUAGES — 56%

Figure 3 - Comparison of the rate of fact-checked misinformation content in English (US), English (UK+Ireland) and in non-English that is unactioned on Facebook

The goal of our study was to estimate the amount of misinformation that Facebook was able to detect on its platform. This is why we included misinformation identified by fact checkers who are Facebook partners but also other reputable fact-checking organisations. But in order to identify the percentage of content Facebook could easily have acted on, having been provided fact checks by its partners, in Figure 4 we also look at the percentage of unlabelled content for which a fact check from a Facebook partner was available. Our data shows that for 66% of the misinformation content that remained unactioned on Facebook, the platform had been provided a fact check from a partner organisation.

FACEBOOK PARTNERS VS. OTHER FACT CHECKERS

PERCENTAGE OF UNLABELLED CONTENT FACT CHECKED BY FACEBOOK FACT CHECKING PARTNERS VS. OTHER FACT CHECKERS

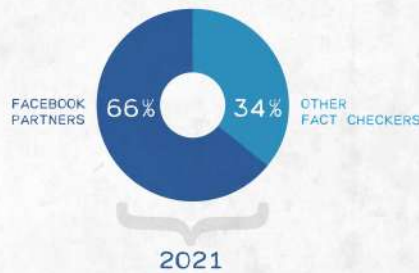FACEBOOK PARTNERS 66% 34% OTHER FACT CHECKERS

2021

Figure 4 - Percentage of unlabelled content that is fact checked by a Facebook partner vs percentage of unlabelled content that is fact checked by reputable fact-checking organisations not part of Facebook's fact-checking network

Next, we focused on the average delay between post publication and Facebook labelling. Unfortunately in 2020 we were not able to record this delay for a significant number of examples, so we only provide it for our 2021 analysis. Avaaz cannot analyse the causes of the delay between when Facebook is notified of a fact check and when it places a label on the content or removes it. Facebook cannot, indeed should not, act before it receives confirmation from a fact-checking partner that a post contains verified misinformation. But the data in Figure 5, though only relating to 26 examples where we were capable of measuring this delay, is indicative of the efficacy of Facebook's system in relation to its actions after receiving a fact check from one of its partners. The exact points of the delay in the system cannot be interrogated without access to further data, which Facebook does not currently provide. What is clear is that this system needs urgent optimisations.

FACEBOOK MEASURE DELAYS

AVERAGE DELAYS BETWEEN POST PUBLICATION AND FACEBOOK MEASURE

2021

DAY 1 DAY 2 DAY 3 DAY 4 DAY 5 DAY 6 DAY 7

28 DAYS

Figure 5 - Delay between fact-checked post publication and Facebook measure (label or removal) in 2021

In Figure 6 we can see that for post delays, too, there is a difference between English and non-English

content, with non-English content on average taking six more days to be labelled or removed. This means non-English speakers on Facebook waited almost one week longer than English speakers to see a label on misinformation or for the content to be removed. As we explain in the recommendations below, Facebook can fix this by being transparent with all users who have been exposed to fact-checked misinformation including through retroactive notifications.
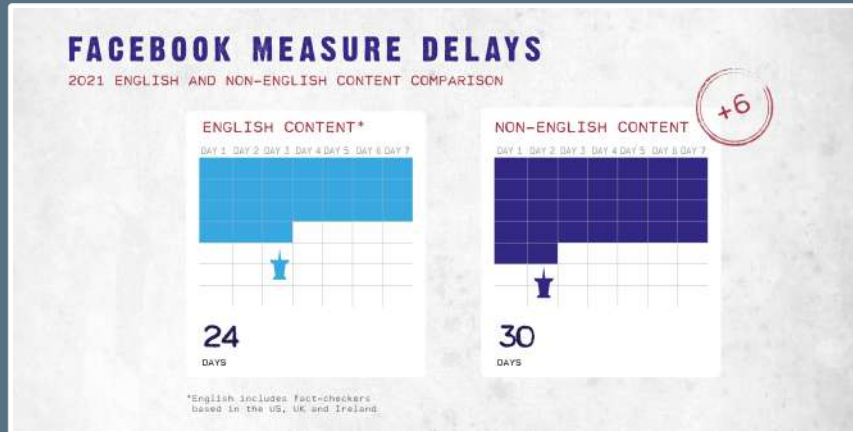

Figure 6 - Delay between publication of post and Facebook labelling for English vs non-English content

Finally, Figure 7 shows the average delay for fact checking to be published after a post's publication, which in 2021 has been nine days. This may be due to a number of variables, including an increased amount of disinformation to fact check or an attempt by fact checkers to focus on older or more viral disinformation on the platform. Fact-checkers are doing their best to keep up with the scale of the problem, under extremely difficult circumstances, yet our findings highlight the need for more capacity on their end. It must also be noted that while Figure 7 relies on a data set of 135 posts, Figures 5 and 6 rely on a different and smaller data set of 26 posts, as this is the number of posts for which we were able to record the exact moment when a label was applied or the post was removed [14] . This means that the nine day delay for fact checking a publication and the 30-day delay on average for labelling it cannot be directly compared, although it suggests that labelling a publication takes much longer than fact checking it.

Figure 7 - Average days for fact checking publications after they are posted

Top COVID-19 Misinformation Narratives and Egregious Falsehoods

The 135 posts analysed in this study are clustered around 23 different narratives. In Figure 8, each circle represents one narrative, and its size is proportional to the number of interactions that narrative collected overall.

The top misinformation narrative we identified (36 posts, 1.4M interactions) is about vaccine side effects, including death, which is very worrying as it could boost vaccine hesitancy right when the world is in the middle of a third COVID-19 wave. False claims 1 and 2 in this section belong to this category.

The second misinformation narrative is about false official measures or warnings (16 posts, 536K interactions), i.e. the WHO has declared that it is not necessary to wear masks [15] . This narrative is dangerous as it erodes trust in official health institutions.

The third biggest narrative is about masks being ineffective or even dangerous (nine posts, 259K interactions). False claims 3 and 4 in this section, and 7 in the following one, belong to this category.
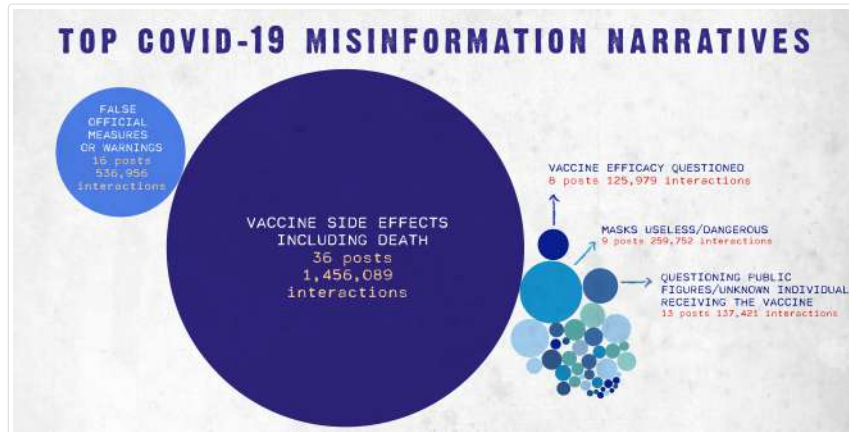

Figure 8 - Top COVID-19 misinformation narratives: Each circle represents one of the narratives we have identified in our data set, with bigger circles representing narratives that have gathered more interactions.

Below, to give a sense of the specific content analysed in this report, we present six false claims [16] that we found during this study. They were selected due to their high interaction rates as well as their potential to cause public harm and further distrust in health authorities.

FALSE CLAIM 1: BILL GATES SAID COVID-19 VACCINE COULD KILL NEARLY A MILLION PEOPLE

This post links to an Italian article claiming that Bill Gates, "[...] during an interview with CNBC, admitted that for every 10,000 people there would be permanent vaccination damage, including 700,000 expected deaths." (translated from Italian, **29,050 interactions**)

Debunked by Facebook's third-party fact checkers Open **and** Facta **and, as** AFP **writes:** "This is false; Gates was talking about vaccine safety and the potential for side effects, and gave a hypothetical figure to illustrate the number of people who could possibly be affected by them worldwide".

Measures: The post was published on January 3, and the article was first published on July 31, 2020. At the end of our investigation, on February 25, no measure had been taken.


Facebook interactions for the article shared in the post.

FALSE CLAIM 2: DOCTOR EXPLAINS THAT VACCINES AGAINST COVID-19 ALTER DNA

This post in Spanish linked to a Facebook video where Dr. Chinda Brandolino states that the "cure", i.e. vaccines against COVID-19, changes your DNA, and negatively affects (male) fertility. **(408,000 interactions)**

Debunked by Facebook's third-party fact checker AFP Factual : "[...] the Argentine "doctor for the truth" Chinda Brandolino assures that vaccines to prevent COVID-19 are "transgenic substances" that "will modify the genes" and will sterilise most men who receive it. But all her warnings are false, according to experts consulted." (translated from Spanish)

Measures: A fact checking article by AFP Factual was available on the post at the time of finding.
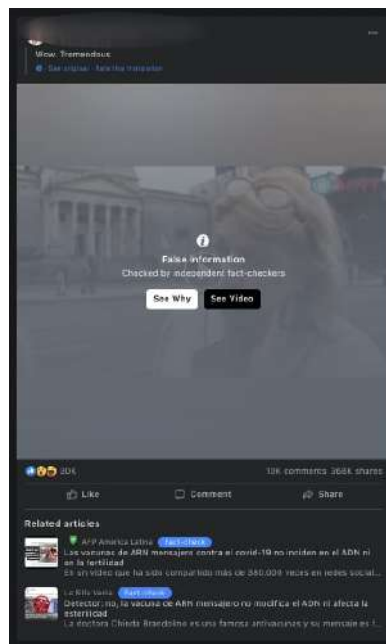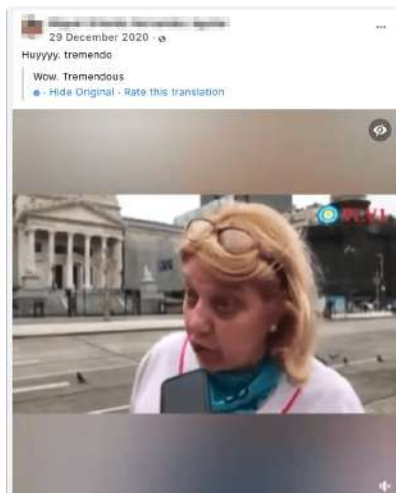

FALSE CLAIM 3: THE REASON WHY FFP2 MASKS BECAME MANDATORY IS BECAUSE THE STATE ORDERED TOO MANY MASKS

This post in German suggests that the reasons behind the mandatory mask wearing is due to the fact that the state has ordered a large number of masks. The post links to a video by the public channel, however it is cherry picking the information provided in the interview. **(226,000 interactions)**

Debunked by Facebook's third-party fact checker Correctiv : "The Facebook user suggests that there is an FFP2 mask requirement because the Federal Ministry of Health has bought too many masks. But that's not true. In Germany there is only an obligation to wear FFP2 masks in Bavaria. The ARD contribution from September 16, 2020, shows that too many masks have been ordered. However, there is no evidence that this was the cause of the introduction of a mask requirement." (translated from German)

Measures: A fact-checking article by Correctiv was available on the post at the time of finding.
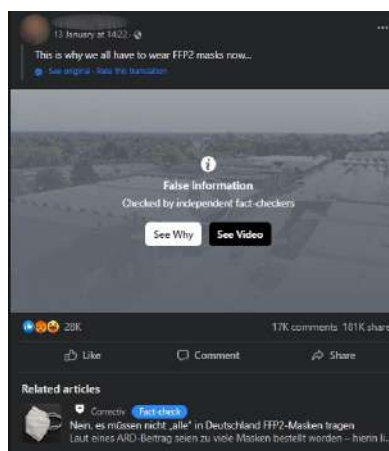
FALSE CLAIM 4: WHO NOW SAYING YOU DO NOT NEED TO WEAR A MASK

This post in English links to a blog claiming that "in a telling admission made on January 22, 2021, the World Health Organization now say [sic] there is no scientific medical reason for any healthy person to wear a mask outside of a hospital. Sadly, our corrupt politicians and mainstream media only relate the bad news." **(27,835 interactions)**

Debunked by Facebook's third-party fact checker Reuters : "False. The WHO has not changed its advice to deem masks unnecessary outside hospital – and has, in fact, strengthened its position on masks being one of many
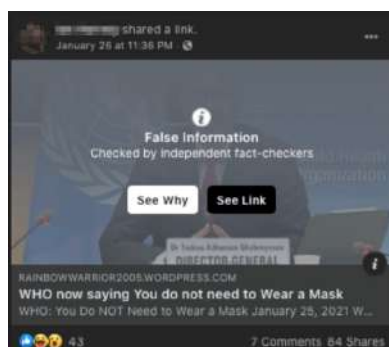
measures that together can help limit transmission of COVID-19."

Measures: A fact checking article by Reuters was available on the post at the time of finding. The post was removed before the end date of this investigation.

The Facebook interaction provided above is based on the article that is shared in the post, and not this specific post.

FALSE CLAIM 5: GREEN BREAST MILK CAN PROTECT CHILDREN FROM COVID-19 INFECTION

This post in Portuguese was published by the Facebook page "Histologia, Fisiologia & Anatomia Humana", with 1.5 million followers. It displays an image comparing two plastic bags filled (allegedly) with breast milk, of which one is visibly green, and claims that "the green colour of the milk of this mother diagnosed with COVID may seem like a bad sign, but on the contrary: It is the result of the antibodies she started to produce to protect her child from a possible infection. Breast milk is so powerful that it is tailored to meet the needs of each baby. That is why breastfeeding, which can be a difficult process for so many mothers, should never be underestimated. Our bodies are capable of a lot - and there is nothing wrong with saying that there is a certain magic in this." (translated from Portuguese, **5,722 interactions**)

Debunked by Facebook's third-party fact checker Polígrafo : Rated False. "Even if the presence of antibodies in breast milk is confirmed, the same does not apply to the greenish colour that appears in the image, which has no relation to the COVID-19 infection. Antibodies are not something that can be seen with the naked eye. Therefore, it is a myth to say that they change the colour of breast milk." (translated from Portuguese) Breast milk can slightly change colour based on the mother's diet, but this does not correlate with antibody production [17] .

Measures: The post was first published on January 25, 2021 and at the end of our investigation, on February 25, no measure had been taken.

FALSE CLAIM 6: ASYMPTOMATIC PEOPLE ARE NOT CONTAGIOUS, ACCORDING TO A CHINESE STUDY

This post in French shares an article that claims asymptomatic people with COVID-19 are not contagious, and that "[...] all this masquerade around 'barrier gestures' is only a political measure to continue the terror." (translated from French, **25,420 interactions**)

Debunked by Facebook's third-party fact checker Le Monde : "False. Asymptomatic patients can still transmit COVID-19. The article distorts the conclusions of a Chinese post-containment study on the population of Wuhan." (translated from French)

Measures: The post was first published on January 1, 2021, and the article remained available on the platform by the end of our investigation, on February 25.

How Facebook is Failing in the "Clone War"

As in the 2020 study, we continued to find debunked misinformation posts that are mutated as "clones" (exact copies) or "variants" (with minor changes to format or narrative) that have succeeded in escaping Facebook's measures. These "clones" and "variants" contribute to exposing millions of users across geographical and language barriers with falsehoods that can potentially cause public harm and further distrust in health authorities.

Here below are two out of the 51 examples of "clones" and "variants" of debunked fake claims that our research team detected; some remain unlabelled and without any measures taken by Facebook.

In both examples below, the post has been shared by individuals, pages and groups who claimed to have found it elsewhere on social media, and as it was not labelled there, they might not be aware of the incorrect nature of the information they posted. This highlights the importance of prompt labelling and notification of false information to users, since without it, the misinformation spreads.

FALSE CLAIM 7: MASKS CAUSE CANCER AND DOZENS OF OTHER DISEASES

This harmful misinformation in Spanish claims that using masks leads to cancer and other diseases [18] . Twelve "clones" and "variants" were found, together reaching **118,288 interactions.**

Debunked by Facebook's third-party fact checker AFP Factual : "Cancer is an abnormal multiplication of cells, therefore it cannot implant or incubate, as the content circulating in networks indicates." [...] Some of the publications that circulate in networks also assure that the masks are a culture of "viruses, bacteria, fungi, parasites" [...] However, these claims do not have scientific support according to specialists previously consulted by AFP Factual." (translated from Spanish)

Measures: Only one [19] "variant" of the posts has been labelled by Facebook. Two examples of posts without any measures taken are here and here .

FALSE CLAIM 8: KAMALA HARRIS WAS NOT REALLY VACCINATED AGAINST COVID-19 - SYRINGE DIDN'T HAVE A NEEDLE

"Clones" and "variants" travel as well across languages. For example, this piece from the US that claimed Vice President-elect Kamala Harris was not really vaccinated against COVID-19 [21] , despite the accompanying video by C-SPAN that all the posts shared. These ten posts have together reached **51,022 interactions** .

Debunked by Facebook's third-party fact checker Reuters : "Social media users have been sharing content online that claims Vice President-elect Kamala Harris's COVID-19 vaccination was faked. This claim is false. [...] There is no evidence Kamala Harris's vaccination was fake or staged."

Measures: This claim was shared in English, Portuguese , French , Italian and German. Here is an example of the English post that was found with a visible fact-check label on it, and examples of posts with no measure on them in Italian and German .

22

Europe's Last Chance to Protect its Citizens from the Infodemic

Social media self-regulation has failed its biggest test to date: This report shows that a year after the start of the COVID-19 pandemic, despite a number of commendable steps to fight the ongoing infodemic, Facebook has failed to improve its ability to detect dangerous COVID-19 disinformation on its platform. This is a key metric in the fight against the infodemic, but the needle has barely moved.

Many of the new policies announced by Facebook, including the platform's expansion of retroactive corrections for a subset of harmful COVID-19 misinformation and its COVID-19 information hub, were small steps in the right direction. But these steps remained piecemeal and fell short of the solutions recommended by experts.

This report uncovers a huge gap between announced initiatives and their implementation, with resources seemingly focused on the United States while major European languages are neglected.

Unfortunately, **current EU initiatives**, including the old Code of Practice on Disinformation and the more recent COVID-19 Disinformation Monitoring Programme have **not even been able to flag the failures highlighted in this report**. Instead they still allow platforms to score themselves on the metrics they choose; and by scoring their own exams, the platforms always succeed.

But **Europe now has a chance to protect its citizens from this infodemic and future infodemics**, right when the continent is being rocked by a third wave of COVID-19. In a few weeks the EU's revised guidelines for a new Code of Practice on Disinformation could, for the very first time, and in combination with the Digital Services Act, introduce real accountability for the harm social media platforms are causing our citizens and our democracies.

But there are three fundamental ingredients that need to be included in the Code of Practice on Disinformation, or it will be doomed to failure again. It needs to demand of platforms and keep them accountable for:

1. **Full transparency toward all users who are exposed to disinformation, including retroactive notifications.** Once again, this report shows that even when labels are applied, they take on average 28 days to be posted, and millions of users who have been exposed in that time will never know they have seen dangerous misinformation. A decade of research on debunking [23] disinformation shows that transparency toward users is one of the most effective tools in fighting it. This means that such a requirement from Europe could help provide reliable health information to tens of millions of its citizens. After the release of our report in 2020, Facebook began to move in this direction, providing retroactive corrections to all users who engaged with what the platform termed "harmful" COVID-19 misinformation. But this is only a small subset of the misinformation on the platform, and Facebook must be regulated to ensure it applies this policy to all misinformation content. The EU can make that the standard.
2. **Detoxifying the algorithm: reduction of the acceleration, caused by the algorithm, of harmful content and systematic misinformers**. Failing to identify a majority of the content in major European languages, as this report shows, also means that Facebook is not able to diminish the acceleration of such content. Instead of being automatically amplified or promoted by the algorithm, repeat disinformers should be downranked. This should be done transparently and include a right to appeal against any demotion. All of this can be done on the basis of fact-checked misinformation and does not require general monitoring. It is worrying that the current discussion around the Code of Practice seems to be focused on vulnerabilities to external manipulation by malign actors, while more and more studies and investigations point to platforms' own algorithms as likely being mainly responsible for the acceleration of misinformation. The EU should push for an independent audit of the role of the platform's algorithms to better understand how to prevent them from acting as misinformation accelerators.
3. **Disclosing the total amount of disinformation present on their platforms and pledging to reduce it over time.** We need to start treating disinformation like we treat CO2. We cannot ban it, but we can keep platforms accountable for reducing it over time to levels that are less toxic to society. Disclosure should be done by the platforms themselves, who should report the disinformation they know about based on key indicators designed by regulators, and also externally, by an independent monitoring body, for a more holistic assessment.

In the immediate term, progress towards these commitments needs to be measured through clear metrics, an independent monitoring body with regulatory experience and the involvement of a wide set of stakeholders.

Avaaz and other civil society organisations are working with experts and regulators to design such policies. Solutions that protect freedom of expression, users' health, and our democracies from the threat of misinformation are available. What is currently needed is the political will from European leaders to ensure that the platforms are regulated and held accountable for the harms they cause to society. This report highlights the urgency of taking action, as the current infodemic in the midst of this pandemic is putting lives at risk.

Centrally, to ensure platforms do not use loopholes to evade their responsibilities, it will be crucial to enforce the new solutions in the Code of Practice on Disinformation be enforced through the Digital Services Act[24].

Methodology and Data Set

The investigative team analysed misinformation content about the coronavirus posted between December 7, 2020, and February 7, 2021, that met the following criteria:

1. Were fact checked by Facebook's third-party fact-checking partners or other reputable fact-checking organisations. [25] [26]
2. Were rated "false" or "misleading," or any rating falling within these categories, according to the tags used by the fact-checking organisations in their fact-check article.
   - The variation of ratings description is quite broad. Here are some examples (please contact us to see the full list we used):
     - Disinformation - Factually Inaccurate - Cherry-Picking - False - Mostly False - Hoax - Incorrect - Misleading - No Evidence - Not True - Wrong
3. Could cause public harm by undermining public health. Avaaz has included content that impacts public health in the areas of:
   - *Preventing disease* : e.g., false information on diseases, epidemics and pandemics and anti-vaccination misinformation.
   - *Prolonging life and promoting health* : e.g., bogus cures and/or encouragement to discontinue recognised medical treatments.
   - *Creating distrust in health institutions, health organisations, medical practice and their recommendations* : e.g., false information implying that clinicians or governments are creating or hiding health risks.
   - *Fearmongering* : health-related misinformation that can induce fear and panic, e.g., misinformation stating that the coronavirus is a human-made bio-weapon being used against certain communities or that Chinese products may contain the virus.

METHODOLOGY FOR MEASURING FACEBOOK LABELLING AND REMOVALS (FIGURES 1, 2, 3, 4, 7)

For the purpose of measuring Facebook's stated claims about its fact-checking efforts, the investigative team analysed a sample of 135 posts about the coronavirus based on the above criteria that were comparable to the sample analysed in our 2020 study. In order to allow a precise comparison with our 2020 study on coronavirus-related misinformation we selected only content that was published in the five languages previously covered in our 2020 report [27] : English, Italian, Spanish, French and Portuguese.

For each of the false and misleading posts and stories sampled based on the above criteria, Avaaz researchers recorded and analysed, using both direct observation and CrowdTangle [28] :

- The total number of interactions it received;
- The total number of views it received in the case of Facebook videos;
- Whether each had a warning label as false or misleading [29] added to it by Facebook [30] ;
- When misinformation posts would receive a fact-check warning label or be removed [31] ; and
- The delay between when the misinformation content was posted and the publication of a fact check by a reputable fact-checking organisation [32] ;

METHODOLOGY FOR MEASURING LABELLING AND REMOVAL DELAYS (FIGURES 5 AND 6)

For the purpose of measuring Facebook's delay between when the original misinformation content was posted and the date Facebook first applied its moderation policies on the post by either flagging the content as misinformation or removing it from the platform, our team analysed a data set of 26 posts. These were all the posts for which our team, accessing them on a daily basis for 43 days, was able to manually document when a label was first applied, or the post was removed.

In order to collect a significant sample, given the difficulty in recording exactly the day when a measure is applied, we considered posts in English, Italian, Spanish, French and Portuguese, but also German.

In our previous discussions with Facebook, we had requested that the platform provide more access to its systems or provide transparency on both the average time between when misinformation is posted, and when a fact-check label is applied or the content is removed, as well as the number of users that view misinformation content before it is labelled. These data points are important as a means of analysing the platform's effectiveness in combating misinformation. We continue to urge Facebook to be more transparent about these numbers more broadly, as it is difficult for researchers to continue to manually conduct such an analysis.

METHODOLOGY FOR IDENTIFYING "CLONES" AND "VARIANTS"

During the research process, our investigative team noticed that posts previously documented were spreading in different languages in an exact, or slightly altered, fashion, and were collecting a large amount of interactions. Our team further investigated seven narratives from our sample of 135 posts, to conduct a dedicated research of the spread of such "clones" and "variants".

We used CrowdTangle [33] to search text from the original post we had documented to identify public shares of the same content - or variations of it - shared by Facebook pages, public groups or verified profiles.

We only included posts when we were able to document at least one occurrence that had been labelled by Facebook but we could also find "clones" or "variants" of the same example that had not been labelled.

With such methodology our team was able to identify a total of 51 posts. The engagement data we estimate for our sample provides some indication of the relative reach of different claims.

GENERAL NOTE ON METHODOLOGY

It is important to note that, while we collect data and compute numbers to the best of our ability, this analysis is not exhaustive as we looked only at a sample of fact-checked misinformation posts in five languages. Moreover, this research is made significantly more challenging because Facebook does not provide investigators with access to the data needed to measure the total response rate, moderation speed, number of fact checks and the amount of users who have seen or been targeted with misinformation.

Nonetheless, Facebook is becoming more cooperative with civil society organisations, and we hope the platform continues this positive trend. We also recognise the hard work of Facebook employees across different sub-teams, who have done their best to push the company to fix the platform's misinformation problem. This report is not an indictment of their personal efforts, but rather highlights the need for much more proactive decisions and solutions implemented by the highest levels of executive power in the company.

This study achieved its purpose by taking a small step towards a better understanding of the scale and scope of the COVID-19 misinformation infodemic on Facebook.

Cooperation across fields, sectors and disciplines is needed more than ever to fight disinformation and misinformation. All social media platforms must become more transparent with their users and with researchers to ensure that the scale of this problem is measured effectively and to help public health officials respond in a more effectual and proportional manner to both the pandemic and the infodemic.

A list of the pieces of misinformation content referenced in this report can be found in the annex.

It is important to note that although fact checks from reputable fact-checking organisations provide a reliable way to identify misinformation content, researchers and fact checkers have a limited window into misinformation spreading in private Facebook groups, on private Facebook profiles and via Facebook messenger.

Similarly, engagement data for Facebook posts analysed in this study are only indicative of wider engagement with, and exposure to, misinformation. Consequently, the findings in this report are likely conservative estimates.

For more information and interviews:
- media@avaaz.org
- **Andrew Legon** (CET timezone) - andrew.legon@avaaz.org / +34 600 820 285

More information about Avaaz's disinformation work: Avaaz is a global democratic movement with more than 66 million members around the world. All funds powering the organisation come from small donations from individual members.

This report is part of an ongoing Avaaz campaign to protect people and democracies from the dangers of disinformation and misinformation on social media. As part of that effort, Avaaz investigations have shed light on how Facebook was a significant catalyst in creating the conditions that swept America down the dark path from election to insurrection ; how Facebook's AI failed American voters ahead of Election Day in October 2020; exposed Facebook's algorithms as a major threat to public health in August 2020; investigated the US-based anti-racism protests where divisive disinformation narratives went viral on Facebook in June 2020; revealed a disinformation network with half a billion views ahead of the European Union elections in 2019; prompted Facebook to take down a network reaching 1.7M people in Spain days before the 2019 national election ; released a report on the fake news reaching millions that fuelled the Yellow Vests crisis in France; exposed a massive disinformation network during the Brazil presidential elections in 2018; revealed the role anti-vaccination misinformation is having on reducing the vaccine rate in Brazil; and released a report on how YouTube was driving millions of people to watch climate misinformation videos .

Avaaz's work on disinformation is rooted in the firm belief that fake news proliferating on social media poses a grave threat to democracy, the health and well-being of communities, and the security of vulnerable people. Avaaz reports openly on its disinformation research so it can alert and educate social media platforms, regulators and the public, and to help society advance smart solutions to defend the integrity of our elections and our democracies. You can find our reports and learn more about our work by visiting: https://secure.avaaz.org/campaign/en/disinfo_hub/ .

Annex

TABLE WITH 10 SIGNIFICANT EXAMPLES OF MISINFORMATION CONTENT REFERENCED IN THIS BRIEF

	Fake News Post	Language	Facebook measure by Feb 25 (None, Labeled, Removed)	Best available interaction	Fact check (linked to article) * = Facebook third-party fact checker
1	Dr. Chinda explains that the "cure", i.e. the vaccine changes your DNA	Spanish	Labeled	408,000	* Factual AFP
2	Asymptomatic people are not contagious	French	None	25,420	* Le Monde
3	The greenish colour of this mother's milk diagnosed with Covid-19 is the result of the antibodies she produces to protect her child from a possible infection	Portuguese	None	5,722	* Poligrafo
4	The reason FFP2 masks became mandatory is because the state ordered too many masks	German	Labeled	226,000	* Correctiv
5	WHO: It is not necessary to wear masks	Italian	Labeled	43,180	* Facta.news
6	Post claiming Bill Gates said there would be 700,000 death from the vaccine and that it is acceptable	Italian	None	29,050	Bufale
7	WHO says you do not need to wear a mask	English	Removed	27,835	* Reuters
8	RNA vaccines cause infertility	French	None	12,500	* Factual AFP
9	Kamala Harris was not really vaccinated against Covid-19 - Syringe didn't have a needle	German	None	32,023	* Correctiv
10	Using masks can lead to cancer and dozens of other diseases	Spanish	None	2,871	* Factual AFP

Endnotes

1. French, Spanish, Portuguese, Italian. ▲
2. When fact-checking sources from the US, UK and Ireland are considered together, 29% of posts were unlabelled. However, if we distinguish between European English and US English, 50% of the UK and Irish fact-checked posts were unlabelled compared to 25% of US posts, meaning the US market is far better served. ▲
3. 20 out of 29 posts containing debunked misinformation. ▲
4. 15 out of 26 posts containing debunked misinformation. ▲
5. 5 out of 10 posts containing debunked misinformation. ▲
6. 5 out of 15 posts containing debunked misinformation. ▲
7. This figure indicates level of labelling of misinformation at the end of each research period. ▲
8. See examples of "clones" and "variants" in the section How Facebook is failing in the "Clone war". ▲
9. Translated from Italian from this article that was shared in the Facebook post. ▲
10. See all 12 posts in the section, How Facebook is failing in the "Clone War". ▲
11. 'Facebook has a blind spot': why Spanish-language misinformation is flourishing, the Guardian, March 2021. ▲
12. French, Spanish, Portuguese, Italian. ▲
13. When fact-checking sources from the US, UK, and Ireland are considered together, 29% of posts were unlabelled. However, if we distinguish between European English and US English, 50% of the UK and Irish fact-checked posts were unlabelled compared to 26% of US posts, meaning the US market is much better served. ▲
14. For more information please refer to the Methodology section. ▲
15. Debunked by fact checker Facta: rated False. "This news is not reflected in any national or international newspaper and in any official press release of the World Health Organization. It is therefore a fictional piece of news." (translated from Italian). ▲
16. In order to ensure a fair comparison with last year we included data from the five languages that were presented in the 2020 study. German was not one of the five initial languages analysed in 2020. In this section, we have nonetheless deliberately chosen to highlight a falsehood in German, as it is a major language and there are nearly 43 million German Facebook users. Data from German examples is not included in the calculations for this study. ▲
17. Yazgan H, Demirdöven M, Yazgan Z, Toraman AR, Gürel A. A mother with green breastmilk due to multivitamin and mineral intake: a case report. Breastfeed Med. 2012;7:310-2. doi:10.1089/bfm.2011.0048. ▲
18. Debunked by Facebook third-party fact checker AFP Factual: "Cancer is an abnormal multiplication of cells, therefore it cannot implant or incubate, as the content circulating in networks indicates." (translated from Spanish). ▲
19. NB: this post is a "variant" of the fact-checked narrative about the danger of masks. It is also the only one we found that had a fact-checking article available under the post. The fact-checking article uses the image (zombie with a mask) that is used in all the other "clones" that we have collected and used as examples in this brief. ▲
20. Data gathered via CrowdTangle Intelligence a public insights tool owned and operated by Facebook and adapted for the needs of this research. ▲
21. Debunked by Facebook third-party fact checkers Reuters and Correctiv, which writes: "A video is shared on Facebook that is supposed to suggest that the future US Vice President Kamala Harris was not really vaccinated against COVID-19 because no needle was to be seen. This is misleading - the needle is clearly visible on higher quality images." (translated from German). ▲
22. Data gathered via CrowdTangle Intelligence a public insights tool owned and operated by Facebook, and adapted for the needs of this research. ▲
23. The Debunking Handbook 2020, George Mason University, Center for Climate Change Communication. ▲
24. Avaaz Position Paper on the Digital Services Act, Disinformation and Freedom of Speech, Feedback to the European Commission, March 31, 2021. ▲
25. Spanish - Maldito Bulo, Newtral, AFP Factual; French - 20 minutes, AFP Factuel, Decodex - Le monde, Les Observateurs de France 24; Italian - Bufale, Butac, Open, Facta; Portuguese - Observador Fact-Check, Polígrafo; English (UK and Ireland) - Full Fact, The Journal.ie; English (USA) - Politifact, Factcheck.org, AFP Fact Check, Lead Stories, AP Fact Check, Reuters Fact Check, USA Today Fact Check, The Dispatch Fact Check, Snopes.com; Health Feedback (NB: HF is registered in France but its reviews are in English). ▲
26. Content fact checked between January 1 and February 25, 2021. ▲
27. How Facebook can Flatten the Curve of the Coronavirus Infodemic, Avaaz, April 2020. ▲
28. Data from CrowdTangle, a public insights tool owned and operated by Facebook. ▲
29. See point 2 of the current methodology for examples of the fact checking ratings used in this study. ▲
30. Examples of warning labels: 1. Fact-checking articles shown as "related articles" below the post; or 2. a gray overlay titled, "False or misleading information checked by independent fact checkers," linking to a fact checking article(s); or 3. A black box titled, "Missing context: Independent fact checkers say this information could mislead people. See why". ▲
31. The Avaaz research team monitored misinformation posts with no warning labels daily between January 13 and February 25, 2021. ▲
32. Fact-checking delays were calculated only for posts that were labelled by Facebook at a later stage. ▲
33. Data from CrowdTangle, a public insights tool owned and operated by Facebook. ▲