

Letter from Rebecca Stimson, Facebook, 26 June 2019

Question 1:

Can and will the 'correct the record' tool use 'linger time' (i.e. dwell time/time users spend on a post) to distinguish between those who do and don't read misinformation to address our concerns that corrections are not being sent to those who haven't shared/engaged with posts but may still have seen them?

As the Committee is aware, since January 2020, we began applying our long-standing misinformation policy to harmful misinformation about COVID-19. In addition we also recently began incorporating a 'correct the record' notification to users who interacted with this kind of harmful COVID-19 misinformation. This includes users who liked, shared, or commented on the violating post. When we send them a notification, we provide a link to the WHO's Mythbusters website, which has a running list of debunked hoaxes about COVID-19.

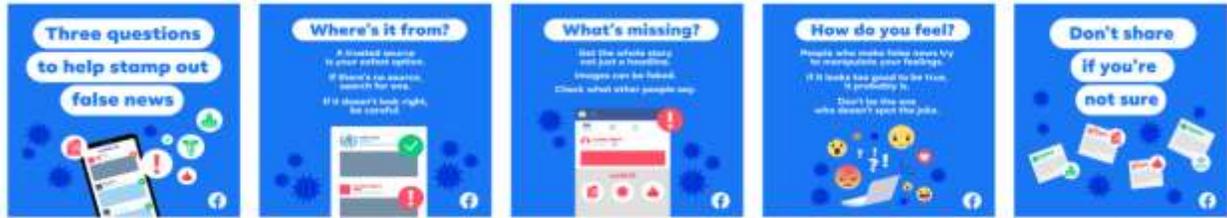
Responding to the specific question about who receives these notifications, we only send 'correct the record' notification to users who liked, commented or shared the harmful misinformation. We do not plan to send this additional information to people who have only seen the content. We believe firstly that this risks potentially drawing unnecessary attention to misinformation to users who in all likelihood only passed by it and secondly it also risks diluting the impact of receiving a notification if it becomes too wide spread and commonplace, which is likely if its sent to everyone who may have seen this kind of content. People who engaged with a piece of content is a higher indication to us that they actually read and potentially absorbed the information contained within it, the priority is therefore to make sure those people are connected with accurate information.

For misinformation about COVID-19 that does not contribute to imminent physical harm, we work with our network of independent third-party fact-checkers. If a user posts or shares misinformation that was rated false by fact-checkers, we send them a notification letting them know that something they posted or shared had been rated by a fact-checker and we link them to the fact-checker's debunking article.

We also want to give people the tools to make informed decisions and judgements about the information they see online and where it comes from. To support this effort, over the coming weeks we'll be rolling out a new campaign in countries across EMEA to educate and inform people about how to detect potential false news.

In consultation with some of our fact-checking partners, including Full Fact, we have developed '[Three Questions To Help Stamp Out False News](#)', which will show up on Facebook through a series of creative adverts, and link out to a dedicated website, www.stampoutfalsenews.com. These adverts will ask people to challenge the information they see on posts by asking themselves the following:

- 1. Where's it from?** *If there's no source, search for one.*
- 2. What's missing?** *Get the whole story, not just the headline.*
- 3. How does it make you feel?** *People who make false news try to manipulate feelings.*



The campaign will roll out in the UK in July. To assess the effectiveness of the campaign in helping people better understand what to trust, we'll follow this campaign with surveys to ask what people have learned. This information will help us create more of these media literacy initiatives in the future.

Question 2:

Can you provide more information on research you have carried out about the effect of de-metification on your platforms in tackling misinformation and disinformation, and separately on the self-esteem and well-being of children and young people?

Responding to the first part of this question - for misinformation that does not violate our community standards but has been rated false by a fact-checker, our approach is to demote the content in News Feed. When we do this, we are dramatically reducing the likelihood that people will come across the content and helps us stem the virality of misinformation. In the month of April, based on 7,500 fact checks on COVID-19 misinformation globally, we applied 50 million warning labels to content on Facebook and when users saw the warning labels, 95% of the time they did not click through to view the false rated content.

Misinformation Research

Measuring false news is hard, because it's hard to define. There's a spectrum of content that different people consider to be 'false news', from hoaxes to opinion pieces to satire, so there is no consensus on how to quantify it. To improve measurement and transparency in the absence of an agreed-upon definition, we're doing a few things.

We're partnering with third-parties, including academics, to help us get additional measurements and are working towards sharing what we learn. For example, we launched a new initiative to help provide independent research about the role of social media in elections, as well as democracy more generally. This commission, called [Social Science One](#), has spent the past year taking proposals for research to measure the volume and effects of misinformation on Facebook

We're providing Social Science One with over [one petabyte of data](#) for this research, which includes a collection of links that people engage with on Facebook. Academics that the commission selects will be able to study these links to better understand the kinds of content that's being shared on Facebook.

In April 2019, our partners at Social Science One and the Social Science Research Council (SSRC) [announced](#) the first grants for independent research. These researchers have access to privacy-protected Facebook data, more than 60 researchers from 30 academic institutions across 11 countries were chosen through a competitive peer review process organized by the SSRC. This externally-validated research will help keep us accountable and track our progress.

We also know that [several studies](#) from academic and journalistic institutions -- using different methodologies-- all find that the overall volume of false news is trending downward since the 2016 election, thanks to the many steps we've taken since then to reduce its spread.

For example, a study about the 2018 midterm elections, conducted by researchers from several American universities, found that fake news exposure fell dramatically from 2016 to 2018, and that Facebook's role in the distribution of misinformation has been dramatically reduced in that time. These results are encouraging and commensurate with our own data, but we know that this is a highly adversarial space and we have more work to do.

Finally, we announced in February this year a \$2 million commitment to support independent social science research on misinformation and polarization related to social communication technologies. This initiative is part of our Content Policy Research Initiative.

Turning to the second part of the question on self-esteem and well-being of children and young people. We understand the Committee's question relates specifically to the removal of 'likes' on Instagram. We want Instagram to be a supportive place where people feel comfortable expressing themselves. We don't want Instagram to be a competition, which is why we've been running a test that removes the total number of likes on people's posts. The idea is to take away the pressure some people might feel, and help them focus on sharing the things they love and feel inspired by – rather than the number of likes they get. This test, which rolled out globally in November 2019, has had positive feedback so far, but is a fundamental change to Instagram and there's still a lot to learn. As this work is incomplete and still evolving we are unfortunately not in a position to share data, but we're continuing to test to understand whether this change can benefit our community's experience on Instagram.

In terms of other research in to social media and well being we shared some of our research insights into social media and well-being last year - link [here](#)

Generally speaking the insights into social media and well being tend to come down to the conclusion that social media is a tool, it's how you use it which dictates whether or not its a positive experience. So some of the insights into facilitating wellbeing include: promote meaningful interaction between people, encourage 1-1 communication, encourage interactions that are effortful (requiring some thoughtfulness).

In response to these findings we now rank News Feed to optimise for meaningful interactions / conversations instead of the metric of time spent on the platform. We do this by predicting which posts you might want to interact with your friends about e.g. posts asking for advice.

We provide people with tools to manage their time on Facebook to see how much time they've spent on the app in the last week, edit their preferences to see content from the people they care most about and daily reminders when they reach the time limits they've set for themselves,

Speaking directly to children and young people:

- Anti-bullying:
 - We have worked with The Diana Award for over seven years supporting their Anti-Bullying Ambassador Programme. As part of this work this year, The Diana Award has created a module focused on how to deal with the pressure to be a

perfect on social media. This is being rolled out throughout the year through the Diana Awards already existing training program in schools throughout the UK.

- We have worked with Childnet for a number of years supporting their schools online safety programme - Digital Leaders. A key element of our 2020/21 partnership involves a series of feedback sessions with young people on the programme where they provide direct and detailed feedback to our teams on our products and services.
- Providing resources to parents and young people:
 - At Instagram, we want the time our community spends on the platform to be positive, inspiring, and intentional. We know from our relationships and work with expert organisations that some people can feel a pressure to look or live a certain way because of social comparison to others. With this in mind we have worked with Internet Matters and Childnet to launch the Pressure to be Perfect toolkit. The toolkit is about recognising that what you see posted by others is just one part of their life - that a single post or video rarely reflects all that is happening behind the scenes. The toolkit is housed on Internet Matters website and is available for download. There is one guide for young people and one designed to support parents.
<https://www.internetmatters.org/resources/helping-young-people-manage-their-onlineidentity/>
 - Unfiltered marketing campaign on Instagram
 - In *Un-Filtered*, a group of teens share their experiences with bullying, hate, and negativity, both on and off-line. Touching on topics of body image, disability, race, and sexuality, *Un-Filtered* is their experiences, in their own words. Watch season one, directed by Jonah Hill, below and on IGTV, now. For resources and to learn more about the issues discussed, scroll down for a list of organizations in your country.
 - <https://www.instagram.com/tv/B2kE72sBd4k/?igshid=8w656bxf76e>
 - Parents guide
 - A parents guide to help navigate Instagram for parents of teenagers
 - <https://about.instagram.com/community/parents>
 - In June this year, we launched a safety guide for school leadership focused on providing guidance for schools on how to safely use Facebook and WhatsApp between teachers as well as parents and the broader community.
 - Your Digital Family:
 - In Summer 2020 in partnership with The Mix, ParentZone, and Connection Coalition, we are due to launch an educational programme focused on providing adults and young people the tools they need to have open, honest, and productive conversations about use of technology and social media.