Full Fact

# Report on the Facebook Third-Party Fact-Checking programme

July 2019 - December 2020

FULL
FACT

# Full Fact is the UK's independent fact checking charity.

# Contents

# Introduction

Full Fact first started working with Facebook on the Third-Party Fact-Checking programme in January 2019. When we joined the programme we committed to reporting regularly on its operation. Our first report, published in July 2019, covered January to June 2019.[1] This second report details our experience from July 2019 to December 2020.

We attempt in these reports to create further transparency around the Third-Party Fact-Checking programme, so that people can reach informed judgements and others can learn from it. We are independent from Facebook and not party to their internal discussions and choices so there are limits to the information we can report on.

This report only covers Full Fact's experience of the programme and does not describe or evaluate Facebook's overall response to the harms that false information causes, or the experience of other fact checkers in the programme.

Our first report set out an overview of how the Third-Party Fact-Checking programme functions, a summary of what work Full Fact undertook in the first six months of the programme, observations from this and recommendations for how the programme could be improved. In this report we give an overview of how the programme has evolved since June 2019, including examining where Facebook has responded to and/or implemented the recommendations we made in our first transparency report.

Broadly, our views are that:

This is a valuable programme. It made a significant difference to our ability to tackle misinformation during the 2019 UK election, and to Facebook's ability to respond. Facebook's global network of fact checking partners meant it had options for responding to misinformation related to the pandemic that other internet companies did not have.

Other internet companies should emulate the Third-Party Fact-Checking programme. In particular, from what we can tell, YouTube stands out as particularly being able to benefit from a similar programme to the Third-Party Fact-Checking programme.

---

1    **fullfact.org/media/uploads/tpfc-q1q2-2019.pdf**

A partnership such as the Third-Party Fact-Checking programme can only be one part of an effective response to misinformation and disinformation. Other decisions the internet companies make are critical and need scrutiny and oversight: from product design, to advertising standards, to rules for user behaviour.

Our two main concerns continue to be transparency and scale. Explaining the programme and its results is Facebook's responsibility. These independent reports from Full Fact seek to add to the information Facebook provides, not act as a substitute.

Most internet companies are trying to use AI to scale fact checking and none is doing so in a transparent way with independent assessment. This is a growing concern.

Tackling misinformation and disinformation in an open society is complex and hard. This report is mainly concerned with the practicalities of the Third-Party Fact-Checking programme and it makes recommendations to Facebook based on our experience. For example, during the 2019 UK election period the programme helped our team to see and respond to claims being made online. However, ambiguity about how to implement the policy that "posts and ads from politicians are generally not subjected to fact-checking" was a practical issue that remained unresolved until new guidance was shared by Facebook in September 2020 ahead of the US election. The UK has thousands of elected politicians including local councillors, and many of them cannot readily be identified as such from their Facebook presence. The guidance on this issue could still be clarified further. We urge Facebook to continue to collaborate with the fact checkers using this programme as they develop further guidance on this and other evolutions of the programme. We offer this report as part of that process, and to keep that dialogue in the open.

All this work needs to be done with open transparent democratic oversight and clear protections for freedom of expression. The Online Safety Bill is a key opportunity for the UK government to demonstrate it can meet this need.

In September 2020 Facebook rolled out a number of new changes to the Third-Party Fact-Checking programme, including new labels for fact checkers to use. In this report we give an overview of these new labels, but do not comment in detail on their effectiveness. We will review the new labels in more detail in the next report. In September 2020 Full Fact also began a pilot of fact checking content received on WhatsApp. The impact of that pilot is out of scope of this report and more information can be found on our website.[2]

---

2   **fullfact.org/blog/2020/sep/full-fact-whatsapp-uk**

## 2020 Recommendations

We make seven recommendations for Facebook and, where appropriate, other internet companies:

- **Recommendation 1:** In the queue, provide data points of number of shares over time, which can be displayed on a graph

- **Recommendation 2:** Invest in better claim matching to identify content

- **Recommendation 3:** Continue to collaborate with fact checkers when developing changes to the programme

- **Recommendation 4:** Continue to review whether users are given the right information when their content is fact checked

- **Recommendation 5:** Provide machine readable data to fact checking partners

- **Recommendation 6:** Share a register of emerging potentially harmful misinformation trends with governments and other relevant official bodies

- **Recommendation 7:** Implement greater transparency around the use of AI in claim matching

## The production of this report

This report was drafted by staff at Full Fact, with input from everybody involved in our work under the Third-Party Fact-Checking programme. The contents are the responsibility of the Chief Executive. They may or may not reflect the views of members of Full Fact's cross-party Board of Trustees and they are not the responsibility of Facebook or any other organisation named in the report. This report has not been shared in advance with other fact checkers who are part of Facebook's Third-Party Fact-Checking programme.

According to the approach we agreed before joining the Third-Party Fact-Checking programme, this report was provided in draft to Facebook on 14 December 2020, with an invitation for Facebook to provide feedback and to respond publicly.

## Facebook's response

Keren Goldshlager, Integrity Partnerships, **Facebook**, said:

"Fact checking has a critical role to play in stopping the spread of misinformation on Facebook and across the broader internet. In the past year, we have grown our global network of fact checking partners to 80 organisations, working in 60 languages, fighting misinformation for critical events like elections and COVID-19. We know our efforts are working. From March to October of 2020, we labeled about 167 million pieces of COVID-19 related Facebook posts, resulting in a 95% drop-off in click-through to the underlying false content.

We welcome Full Fact's reporting on the state of online misinformation and are proud of their observation that ours is 'the most effective response of any internet company so far.' We look forward to improving our work even more – with Full Fact and all our fact checking partners – in the months and years to come."

## Operating guidelines

Our **operating guidelines** have not changed since the 2019 Third-Party Fact-Checking report.

## Funding

Between July 2019 and November 2020, Full Fact received £312,507 for its work on the Third Party Fact Checking programme.

The amount of money that Full Fact is entitled to depends on the amount of fact checking done under the programme.

# Progress on the 2019 recommendations

Our 2019 Third-Party Fact-Checking report made 10 recommendations to Facebook.

Facebook responded to that report stating:

"We are encouraged that many of the recommendations in the report are being actively pursued by our teams as part of continued dialogue with our partners, and we know there's always room to improve… We look forward to continued collaboration with Full Fact and our more than 50 global fact-checking partners."

We welcomed Facebook's commitment to working with us, and other fact checkers, to improve the programme. Some of the recommendations made in 2019, such as to include Instagram in the programme, were implemented swiftly following the publication of our report.

On 11 August 2020 Facebook publicly announced[3] a number of new updates to the Third-Party Fact-Checking labels available to fact checkers, which was made available to UK Third-Party Fact-Checking partners on 24 September. We are pleased that this includes many of the recommendations that we called for in 2019. However, we are disappointed that it has taken over a year for these changes to be put in place.

Below we review Facebook's response to each recommendation in turn.

**Recommendation 1: Continue developing tools that can better identify potentially harmful false content, including repeated posts**

Within its fact checking product, Facebook provides fact checkers with a queue of publicly-shared posts that have been identified as potentially needing fact checking from both its own systems and from user reports. From our experience the majority of the queue's contents are posts that are not fact checkable, usually because they don't contain actual claims or contain claims that are not relevant to the UK and therefore fall outwith our editorial remit. Our 2019 report highlighted our view that there must be more potentially harmful false content than we were able to see, and this is still the case.

---

3   **facebook.com/journalismproject/programs/third-party-fact-checking/new-ratings**

Initially, Facebook slightly improved the queue by providing a 'Related Items' column under posts we had already rated. It was a step in the right direction but had limited use as often no related articles would appear.

As of September 2020, there is a 'Suggested Matches' tab directly in the tool. These matches are much improved in terms of the claim identified, and often help us to find duplicates. There is certainly still some room for improvement, but we recognise that we are benefitting from developing technology and we welcome this. As experts in artificial intelligence in this field we know how hard it is to get right and the risks from imperfect technology. The results we have seen so far emphasise the need for independent scrutiny of the use of AI to tackle harmful false information online by all internet companies.

### Recommendation 2: Provide more data on shares over time for flagged content

Facebook has introduced the option to display shares over the last 24 hours and the last 120 days for posts on the queue. This is helpful in giving fact checkers an idea of whether something is currently going viral, or if the sharing of it has slowed down.

But this data can't be used to get the full picture of where the post is in its journey of going viral. While we understand that some of this information is available on CrowdTangle, a public insights tool from Facebook, we recommend introducing data points of number of shares over time, which can be displayed on a graph, directly within the tool used to rate content.

### Recommendation 3: Add a 'Mixture' rating which does not reduce the reach of content

Facebook renamed the 'Mixture' rating to 'Partly False' based on user research which it says shows that the mixture label confused users. Previously content labelled as 'Mixture' had its distribution on newsfeeds reduced, usually with a grey overlay if the misinformation is an image or video. These posts also had their distribution reduced less than posts rated as 'False'. This has not changed with the new 'Partly False' label.

Our recommendation was to have a rating that indicated some of the content may be false, but that did not reduce the distribution on newsfeeds. Although Facebook introduced a 'Missing Context' rating, which does not downrate content on newsfeeds, this rating isn't for explicitly false information, it's for content that's potentially misleading without context. The change of language from 'Mixture' to 'Partly False' means that in some cases, this newer rating can seem too strong if a post contains a combination of true and false claims. 'Missing Context' may

also not be appropriate if the claim is a mixture of true and false, and context is not the issue.

### Recommendation 4: Add an 'Unsubstantiated' rating

Facebook has not added such a rating. We repeat that this would be beneficial in cases where we cannot definitively say something is false, but equally can find no evidence that it is correct. Facebook states that its new 'Missing context' rating should also apply to "claims stated as fact that are unproven".[4]

### Recommendation 5: Add a 'More context needed' rating

We were pleased to see Facebook introduce a 'Missing Context' label in September 2020.

### Recommendation 6: Add a rating for humorous posts other than satire or pranks

In October 2019 Facebook removed the 'Satire' rating, therefore meaning that for just under a year there was no way to label content originally made as a joke but which had been misconstrued. During that time we were unable to stop the spread of things we previously[5] would have been able to,[6,7,8] including this image about removing a "tracker" from a wheel,[9] or where to put antifreeze in an engine.[10]

Facebook has now reinstated the 'Satire' rating, in September 2020, based on feedback from its fact checking partners.

### Recommendation 7: Develop clearer guidance on how to differentiate between several claims within a single post

Facebook has changed the process of rating content. While the entire post is given one rating, fact-checkers now also have the option to indicate which components/details within the post are false, for example just the text, just the photo, the text and photo together, or the text and photo separately. Facebook then uses these details to find identical content. This gives us some way to show images have been taken out of context, but it does not resolve this recommendation.

---

4    **facebook.com/business/help/341102040382165**

5    **fullfact.org/online/wind-turbine-explosion**

6    **fullfact.org/online/wind-turbine-explosion**

7    **fullfact.org/online/picture-mum-snorting-drugs**

8    **fullfact.org/online/freddos-vs-inflation**

9    **eatliver.com/rfid-chip**

10   **facebook.com/465404940859107/photos/a.465407467525521/551148355618098**

Fact checkers still have a problem with how to rate pieces of content that contain multiple claims, where they are a combination of true and false. For example, it has not always been clear if the new 'Partly False' label (as discussed under Recommendation 3 above) means that the claims within a post have to be half true and half false. Facebook's help centre[11] now says that one definition of 'Partly False' is "A mix of true and false key claims, where the false claims do not predominate."

It is unclear what action fact checkers should take if the post contains a large number of claims that are a mixture of true, false and opinions, perhaps in an even split. At the moment Full Fact is dealing with this on a case by case basis.

**Recommendation 8: Share more data with fact checkers about the reach of our fact checks**

Data on how many people are seeing fact checks via Third-Party Fact-Checking and, more importantly, whether seeing these checks is changing their behaviour, has still not been made available to fact checkers on the programme. There is only so much individual tracking that can be done by fact checkers themselves, for example using website analytics.

Full Fact was given an 'activity summary' by Facebook for Q1 and Q2 of 2019 which contained general information about how many people had got notifications because of our work, and how many duplicate pieces of content they identified. Before October 2020 we did not receive any summaries, and remain unclear on whether similar summaries were made available to other fact checkers.

However, late in October 2020 Facebook rolled out a new feature which provides a small amount of information about the reach of fact checks done through the programme to individual partner organisations.

In order to evaluate the success of the programme, fact checkers need country specific machine readable data to understand how different groups react to seeing fact checks; the decisions people take and whether this changes according to the rating shown to users; how many users have seen and interacted with fact checks from specific partners; and how user and publisher behaviours might change in the long term. This should be provided at regular, agreed intervals.

---

11  **facebook.com/business/help/341102040382165**

**Recommendation 9: The Third-Party Fact-Checking programme should expand to fully include Instagram content**

We are pleased that Facebook has expanded the programme to include Instagram.

Previously, ratings we applied to Facebook images could affect identical pictures on Instagram. False pictures we rated would also be blocked from the Discover page and hashtags on Instagram. From the end of 2019, Instagram posts appeared in the queue Facebook provides us. A grey overlay now appears on Instagram posts rated 'False', even if the false post is shared in stories or as a direct message. Facebook's rules on politicians and their adverts not being subject to fact checking also applies to their Instagram posts.

**Recommendation 10: Be explicit about plans for machine learning**

Facebook has not effectively taken this recommendation forward. We recognise that Facebook continues to explore the use of a range of AI projects across its products and it has provided more information over the past year about when it uses machine learning.[12] This is welcome.

However, the key question for Facebook and all other internet companies is how they use machine learning, which is complex and can have unintended consequences. This work cannot be done in isolation. Substantial effort is needed to provide clarity on the definitions and the limits of the definitions they use, as well as the data and the limits of any data that powers algorithmic decision making. As well as definitions and data, the product choices that are made or informed by any algorithms need to effectively communicate its limits to users. Analysis and evaluation of misinterpretations needs to be made available, and must feed back into the design process. Without these steps, work in this area is inadequate at best.

Full Fact believes that greater transparency in these choices is the only responsible way to develop this space. We have long advocated for internet companies to commission and publish shared training and evaluation datasets, to continually test for unintended consequences of the use of AI and technology. We are also aware that the one-size-fits all model does not work for a global service that operates across many languages, cultures and societies. Algorithmic choices must reflect these subtleties or effectively understand and communicate algorithms' limits. We would welcome Facebook taking the industry lead in this space and we are prepared to work with them on it.

---

12   For example in these blog posts: **ai.facebook.com/blog/heres-how-were-using-ai-to-help-detect-misin-formation**, **ai.facebook.com/blog/using-ai-to-detect-covid-19-misinformation-and-exploitative-content**

# Other changes since July 2019

A summary of how the Third-Party Fact-Checking programme works is detailed in our July 2019 report.[13]

Facebook has made changes to the programme since then. We summarise the most impactful below.

## Ratings

Facebook has made a number of changes to the label ratings available to fact checkers to apply to content:

- Up to October 2019 the labels available were: **False**, **Mixture**, **False Headline**, **True**, **Satire** and **Opinion.**

- From October 2019 to August 2020 the labels were: **False**, **Partly False**, **False Headline** and **True.**

- From September 2020 onwards the labels are: **False**, **Altered**, **Partly False**, **Missing Context**, **Satire** and **True**

### 'Mixture'

The 'Mixture' rating was replaced with 'Partly False'. Facebook said this change was based on user research, as 'Mixture' had not been well understood by users. Facebook said the underlying definition would remain the same. Content previously rated 'Mixture' now shows up as 'Partly False' on Facebook.

Although we agree that the 'Mixture' rating was difficult for users to understand, the 'Partly False' rating comes across as a completely different category with a different definition. As outlined in the first section, this means we would not apply the 'Partly False' rating to all content we had previously rated as 'Mixture'. There are cases where content was genuinely a mixture of true and false (often posts with several claims) and 'Partly False' is, to us, a completely different rating.

### 'Satire' and 'Opinion'

Facebook removed the 'Satire' rating but then reintroduced it in September 2020. The 'Opinion' rating was also removed, which was previously a rating that didn't

---

13  **fullfact.org/media/uploads/tpfc-q1q2-2019.pdf**

demote the content it was attached to. Although genuine opinions were always exempt from the Third-Party Fact-Checking, there was a rating with that name.

## 'Altered' and 'Missing Context'

In September 2020, Facebook introduced an 'Altered' rating, for pictures, video or audio that has been "edited or synthesised beyond adjustments for clarity or quality, in ways that could mislead people".[14] This content would be given a grey overlay, like false posts, and users would be notified if they had previously shared it, and if they tried to share it.
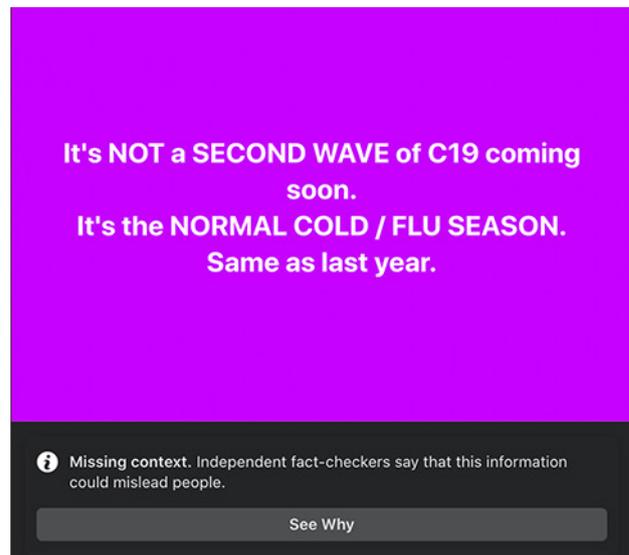


Facebook also introduced the 'Missing Context' rating in September 2020 for content that "may mislead without additional context". By their definition, this would cover content that includes claims stated as fact that are unproven, for example, certain medicines which have the potential to treat Covid-19, but haven't been proven to do so. Content rated this way has a label but does not have the full visual overlay, or have reduced reach on news feeds.

As stated above, we will review the effectiveness of these new labels in our next report.
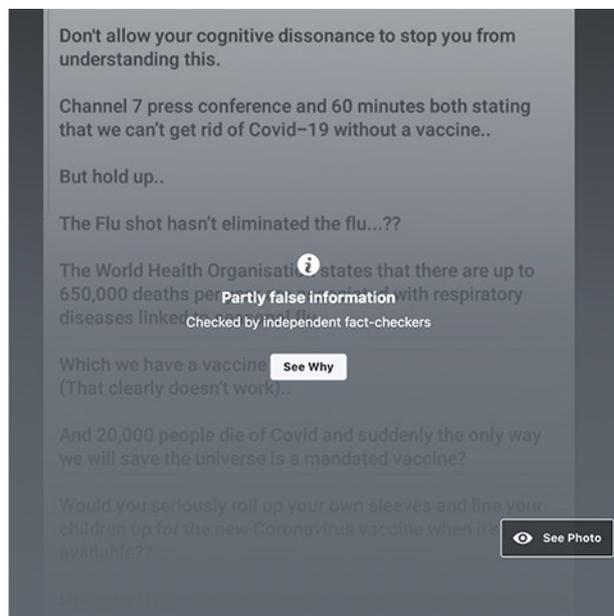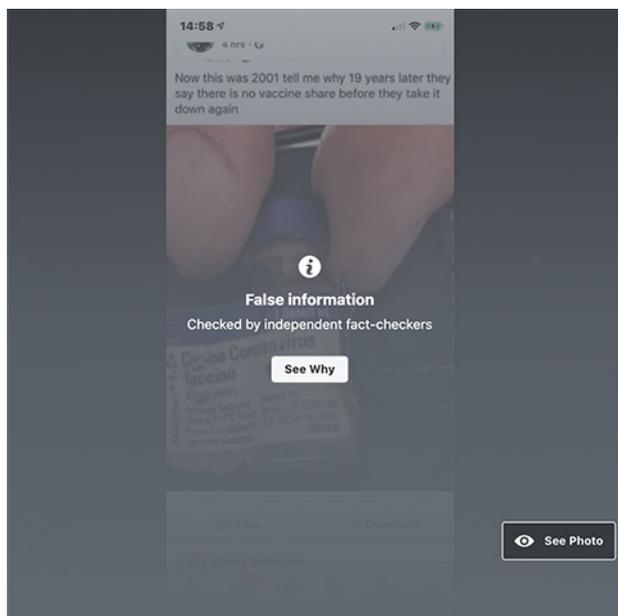
## Overlays

In October 2019, Facebook said:

"Content across Facebook and Instagram that has been rated false or partly false by a Third-Party Fact-Checker will start to be more prominently labeled so that

---

14  **facebook.com/business/help/341102040382165**

people can better decide for themselves what to read, trust and share. The labels below will be shown on top of false and partly false photos and videos, including on top of Stories content on Instagram, and will link out to the assessment from the fact-checker. Perhaps the most obvious change to the Third-Party Fact-Checking programme is how content rated false now appears to users."[15]



A grey overlay now appears over images, and videos and articles on Facebook rated 'False'. When the 'Altered' rating was introduced in September 2020, content rated as such also had a grey overlay."

If the content of the post is just a text-based status, no overlay appears.

## Instagram

Facebook introduced Instagram posts to the queue towards the end of 2019. Previously, if we had rated pictures on Facebook via Third-Party Fact-Checking as false, identical pictures on Instagram would be hidden from the Discover page and hashtags.

From December 2019, Instagram posts have also appeared in the queue, and can be rated like any other piece of content. Instagram TV (IGTV) posts and the recently launched Reels can also appear in the queue. Fact checkers cannot rate content posted to Stories unless it is made into an Instagram grid post, however if a rated post is embedded as an Instagram story, Facebook will show the corresponding label.

---

15  **newsroom.fb.com/news/2019/10/update-on-election-integrity-efforts**

The data provided to fact checkers via the queue about these posts is different to the data provided about Facebook posts. Shares are not reported, which is perhaps more understandable given that Instagram is less share-focused. Views are provided for the past 24 hours and the past 120 days.

Posts rated 'False' also have a grey overlay, similar to Facebook posts, with the option to read more.[16] This overlay also appears if the post is shared via direct message between users, as well as in Stories.

## Politicians' speech

How political speech should be treated by Facebook or other internet companies raises questions of principle and practicality. Reasonable people can and do disagree about these principles. This section is intended to inform these discussions by explaining the practicalities.

For our part Full Fact continues to believe that more choices about the responsibilities of internet companies should be made through open democratic transparent debate. To their credit, Facebook has said the same.[17] The UK parliament has yet to debate these issues substantively, despite the important work done by some Select Committees.

In September 2019, Facebook's VP of Global Affairs and Communications Nick Clegg made a speech in Washington DC during which he discussed political speech on the platform. He repeated the fact that politicians are exempt from the Third-Party Fact-Checking programme, which we had been aware of since the start of the programme.

During the speech he said:

"We do not submit speech by politicians to our independent fact-checkers, and we generally allow it on the platform even when it would otherwise breach our normal content rules. Of course, there are exceptions. Broadly speaking they are two-fold: where speech endangers people; and where we take money, which is why we have more stringent rules on advertising than we do for ordinary speech and rhetoric."[18] Facebook has made clear that this includes the words a politician says as well as photo, video, or other content that is clearly labeled as created by the politician or their campaign.[19]

---

16 **instagram.com/p/B6-QWaDAz5p**

17 **about.fb.com/news/2020/02/online-content-regulation**

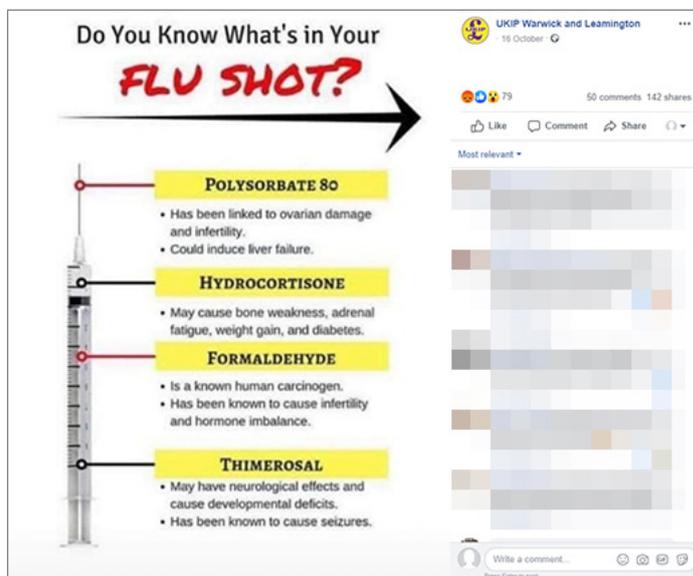18 **about.fb.com/news/2019/09/elections-and-political-speech**

19 **facebook.com/business/help/315131736305613?id=673052479947730**

In a blog post accompanying the speech, he repeated that politicians are exempt from fact-checking, but that "when a politician shares previously debunked content including links, videos and photos, we plan to demote that content, display related information from fact-checkers, and reject its inclusion in advertisements."[20]

Facebook policy includes what it calls a newsworthiness exemption[21] - if someone "makes a statement or shares a post which breaks our community standards we will still allow it on our platform if we believe the public interest in seeing it outweighs the risk of harm."[22]

Mr Clegg continued in the blog post: "Today, I announced that from now on we will treat speech from politicians as newsworthy content that should, as a general rule, be seen and heard." But he added that this would not apply to ads and that "if someone chooses to post an ad on Facebook, they must still fall within our Community Standards and our advertising policies."

In response to criticism of inaccurate content from high profile individuals remaining on the platform, in June 2020 Facebook announced that it would start labelling content that is covered by its newsworthy exemption. It will also give users a warning that content might break community standards when they attempt to share it. In the announcement, Mark Zuckerberg reiterated that "there is no newsworthiness exemption for content that incites violence or suppresses voting. Even if a politician or government official says it".[23]



---

20  **about.fb.com/news/2019/09/elections-and-political-speech**

21  **about.fb.com/news/2016/10/input-from-community-and-partners-on-our-community-standards**

22  **about.fb.com/news/2019/09/elections-and-political-speech**

23  **m.facebook.com/story.php?story_fbid=10112048980882521&id=4**

There are cases where Facebook's definition of political speech would mean we would be prevented from fact checking posts that could cause harm. For example, in October 2019 a regional UKIP Facebook page shared an image containing misinformation about vaccines.[24]

Under the terms of the Third-Party Fact-Checking programme, we were not able to give this content a rating, as Facebook defines it as political speech, despite its obvious potential for harm. It is unclear whether this page is affiliated with the central party, and the party does not, at the time of writing, seem to have any policies against vaccinations.

In mid-March 2020, as lockdown in the UK began, we also saw some local councillors share an out of context image of worshippers outside of a mosque,[25] which was taken before lockdown began. We did not attach the accompanying false rating to content from these figures.

Facebook do say that their Community Standards apply to politicians' speech. This means that they can remove misinformation "which can contribute to imminent physical harm or violence", including harmful misinformation about Covid-19 and more recently Covid-19 vaccine misinfo. This is outside the scope of the Third-Party Fact-Checking programme and is at Facebook's discretion.

In August 2020 Facebook updated their policy to clarify that a "politician" is defined as "candidates running for office, current office holders – and, by extension, many of their cabinet appointees – along with political parties and their leaders". This also clarifies that former candidates for office or former officials continue to be covered by our Third-Party Fact-Checking programme".[26]

Beyond this definition, Facebook has taken the view that who qualifies under the political exemption is best determined by local fact checkers who understand their own political environments, asking "fact-checkers to use their expertise and judgment to determine whether an individual is a politician".[27] However, we believe that there is room for further guidance to ensure consistency internationally.

In evaluating all this, Full Fact's starting point is freedom of expression, the freedom to impart and receive information, and the need for proportionate responses to specific harms within that.

---

24  **facebook.com/UKIPWarwickandLeamington/photos/a.381582272026313/1251844571666741**

25  **fullfact.org/online/coronavirus-mosque-shut-lockdown**

26  **facebook.com/business/help/315131736305613?id=673052479947730**

27  **facebook.com/business/help/315131736305613?id=673052479947730**

As this debate proceeds, we ask people to reflect on whether rules that privilege the political speech of office holders and candidates over other citizens in a democracy succeed in either protecting freedom of expression, or allowing proportionate responses to real harms caused by their speech.

Again we emphasise that there is no perfect answer to the trade-offs here, and certainly none that will apply equally well in the different democracies of the world. These decisions are being made by US-based internet companies and not locally in open transparent democratic processes because of the failure of democratic governments and parliaments to act.

## Opinion and science

Some campaigners against climate change have claimed that the Third-Party Fact-Checking programme has a loophole that allows claims from climate change sceptics to go unchecked, as opinion claims are out of scope of the Third-Party Fact-Checking programme. In at least one particular case some misleading claims about climate science have been ruled ineligible for fact-checking because Facebook has classified them as opinions.[28]

We have not had any issue around this boundary or Facebook's approach to it in our work on the programme and we will report it in these transparency reports if we ever do. Again, this report does not speak for any other fact checker involved in the Third-Party Fact-Checking programme.

We have sought and received assurances from Facebook that they will not take any action which alters the effect of any fact check provided by Full Fact to Facebook without informing us.

However, it has always been the case that you can't fact check an opinion on Facebook or anywhere else, as we discussed in our first transparency report. The line between fact, opinion, and judgement has always been contested, since long before the internet, and long before the climate change debate.

For example, the UK, Cl. 1(iv) of the Editors' Code of Practice[29] says that: "The press, while free to editorialise and campaign, must distinguish clearly between comment, conjecture and fact." Similarly, point four of the National Union of Journalists Code of Conduct[30] says that "a journalist... Differentiates between fact and opinion."

---

28  **nytimes.com/2020/07/14/climate/climate-facebook-fact-checking.html**

29  **ipso.co.uk/editors-code-of-practice**

30  **nuj.org.uk/about/nuj-code**

The boundary between fact, judgement, and opinion can be blurred by experts as well as journalists and anybody else. We are conscious of the risk to freedom of expression from attempts to counter harmful false information overreaching and it is important that scientists and other experts in public debate recognise those risks too. This needs further open and transparent discussion in the context of the Third-Party Fact-Checking programme.

## Appeals

Authors of posts that have been fact checked can contact fact checkers via email if they believe they have corrected the rated content, or want to dispute the fact checker's rating.[31]

Each fact checker has its own specific email address to deal with these queries.

In our experience, this process is not communicated well to authors of posts. Almost all of the emails Full Fact receives to its appeals email address are from people whose content has been rated by other fact checkers, rather than Full Fact. Some assume because their post is in English, or because they are based in the UK, that we have fact checked it, which is not always the case.

They almost always do not know how to find out which fact checker has rated their content, how to contact that fact checker, and sometimes they do not know which of their posts have been fact checked. We understand from Facebook that every Page admin has access to a Page Quality tab that shows active fact checks, including a link to the content and the corresponding fact check article, and that Page and domain owners receive notifications when content has been fact checked. But in our experience this information is not being understood by those who wish to appeal.

Facebook asks that we try to acknowledge emails of this nature within 24 hours, but it takes work to establish whether or not these emails are meant for us, especially as it is not always clear what has happened or what changes could be made to the post. A series of relatively small tweaks would improve this considerably:

- Make it clear to publishers which fact checkers have rated their post, and give them the contact details of that fact checker automatically.

- Make it clear to publishers which of their posts has been rated.

- Make it clear to publishers they need to include a link or screenshot

---

31  **facebook.com/business/help/182222309230722**

of the post they are referring to when making an appeal.

In early 2020 Facebook provided a template for users to use with which they can email fact checkers.[32] We will assess whether this has had a meaningful impact in our next report.

Facebook has also announced the first members of the new Oversight Board.[33] Facebook has stated that in the future, people who use Facebook will have the ability to request a review of the other enforcement actions. For example, this could potentially include content rated false by Third-Party Fact-Checkers on the basis that the content was not eligible for fact checking. We await more detail and engagement on how this will function.

## Changes to workflow

Facebook's product that fact checkers use to submit their ratings has also been changed for the better. From September 2020, fact checkers attach the rating to a post, and also tell the system what part of the post is that rating: whether it's the text of the status, the photo or video itself, the text contained in the image, or all of this together. This was to improve Facebook's ability to find identical content to rate, using our fact checks.

Facebook are also using this extra information to surface similar pieces of content, suggesting additional content to fact checkers to also rate. The quality of these surfaced related items has notably improved in recent months, and this is now a valuable part of the tool.

## Collaboration

We appreciate that Facebook has been open and collaborative with us on specific pieces of research conducted, including a collaborative project testing the impact of different headlines on user understanding and seeking our advice on evaluation of the joint media literacy campaign Stamp Out False News[34] that launched in July 2020.

There is scope for significantly more research and collaboration with external experts. Online misinformation is a new area of inquiry, and fact checkers need field research to greater understand the impacts of ratings, and their various formats, on different users' beliefs and behaviours. It would be beneficial to see Facebook testing the impact of different messaging on posts which have been fact checked, to understand what changes behaviours.

---

32  **facebook.com/business/help/997484867366026**

33  **about.fb.com/news/2020/05/welcoming-the-oversight-board**

34  **stampoutfalsenews.com**

Facebook should continue to work with fact checkers to understand the most effective way to prevent misinformation from spreading on their platform.

# Observations from Full Fact's work since July 2019

Between July 2019 and 14 December 2020 Full Fact has published over 420 fact checks on posts as part of the Third-Party Fact-Checking programme.

## November 2019 UK General Election

The UK General Election was announced at the end of October 2019, with official campaigning beginning a week into November.

The Third-Party Fact-Checking programme helped us to see content related to the election. For example, we checked claims suggesting a viral image of a boy sleeping on a hospital floor had been faked. We rated this 'False'.[35] We attached our reference article to 71 pieces of content on the queue on 10 December, and Facebook told us on 12 December that the rating had been applied to 971 instances of the claim on the platform. We're proud of being able to have this impact on a debate that dominated the news cycle for several days.

However, as outlined above in "Politicians' Speech", Facebook has a standing policy that states: "posts and ads from politicians are generally not subjected to fact-checking."[36]

We asked Facebook multiple times over the course of the election period to clarify the definition of a politician: did it include anyone running for any seat? What about local councillors, or people who were known publicly as politicians but who weren't currently in office? Or those in the running to become Police Commissioners?

Previously, we had not been fact checking direct quotes from anyone in elected office, or the output of registered political parties or official campaigns, via the programme. But we did not receive answers to the questions we posed to the Facebook team. Facebook did not provide updated guidance on the definition of a politician until September 2020 in the context of the US election. This guidance still does not answer all of the questions we posed in 2019.

---

35  **fullfact.org/online/LGI-photo-boy-facebook**

36  **web.archive.org/web/20200104205448/https://www.facebook.com/help/publisher/182222309230722**

At one point, Facebook asked us to remove a rating from a piece of content because it had been posted by a local councillor, and therefore violated the policy. We did remove this rating.

It is currently impossible for fact checkers to determine in all cases whether a user is a political figure or not. With over 10,000 councillors, MPs and peers in the UK alone it is no small task. In the UK this data sits between the Electoral Commission, local councils and non-profits like Democracy Club and mySociety.

There are other complexities that must be considered. For example, individuals may be using personal accounts that don't publicly disclose their roles or they may not be verified or badged appropriately. There is also no guidance on how to treat politicians who may be considered inactive, for example some councillors or MPs who may be on maternity leave. As discussed above, Facebook's general position is that it prefers to leave these judgements to the expert fact checkers it works with. We believe that is compatible with further guidance to ensure consistency internationally.

## Edited Sir Keir Starmer video

During the election, the Conservative party edited footage of Sir Keir Starmer MP being interviewed to make it look like he did not respond to a question on Brexit.[37]

We publicly called this irresponsible.[38] The Conservative party did not delete their Facebook post[39] containing the edited video.

We did not rate the original posts on Facebook via the Third-Party Fact-Checking product, as it would not have been in scope, having been posted by a registered political party. However, we did initially apply a rating to a number of other versions of the video, reposted on Facebook by non-politicians and pages that weren't political parties. Facebook asked us to remove these ratings on those videos, which we did. We also removed the text at the bottom of our article[40] that gave it a rating, as all of our fact checks on content identified through the programme do.

We were surprised by Facebook's decision. The video was, by its very nature, not direct and unedited speech, as per the guidelines at the time. Although the content

---

37  **fullfact.org/news/keir-starmer-gmb**

38  **twitter.com/fullfact/status/1191784713997946885?lang=en**

39  **facebook.com/watch/?v=399805414261833&external_log_id=cdd82f33354b52f80e09a97dbda9e-cb9&q=conservatives%20brexit%20minister%20can%27t%20or%20won%27t**

40  **fullfact.org/news/keir-starmer-gmb**

was originally edited together by the Conservative party, the video itself didn't have any Conservative branding on it. The video was still misleading without the accompanying text in the Conservatives' tweet and Facebook post.

This is another example of where greater discussion is needed on how fact checkers should interpret and enforce the political speech policy. As above, we had previously sought further guidance and advice for how the rules should be interpreted during the election campaign but did not receive clarification on the points we raised. This is urgently needed before the next political event, which in the UK will be in May 2021.

## Covid-19 infodemic

Since the outbreak of Covid-19 in early 2020, we have seen a significant amount of false claims and exaggerations originating and spreading on social media.

The Third-Party Fact-Checking product has proved to be a valuable tool when it comes to monitoring the types of misinformation spreading about the virus, including dangerous misinformation about cures, preventative measures and the origins of the disease. Facebook have taken a number of additional measures to counter Covid-19 related misinformation, and have published information on this on their blog.[41] This has included providing Full Fact with additional funding to hire a health misinformation Fellow.[42]

We have seen a huge amount of misinformation related to Covid-19, including claims about vaccines such as claims about previous vaccines,[43] and whether vaccines can be forced on members of the public.

The fact that when the disease was first named, it was referred to as simply 'coronavirus' certainly fueled some of the early misinformation we checked. A picture of a coronavirus vaccine for dogs[44] was used to peddle the myth that Covid-19 is not a new disease.

Long copied and pasted chainmail-like posts have been common.[45,46] A post on Facebook claiming to be from a user's uncle that was shared over 300,000 times is a good example of a mixture of accurate and inaccurate claims. While the post

41  **about.fb.com/news/2020/10/coronavirus**

42  **facebook.com/journalismproject/programs/third-party-fact-checking/global-health-fellowship**

43  **fullfact.org/online/coronavirus-vaccine-compared-to-flu-vaccine**

44  **fullfact.org/online/dog-vaccine-coronavirus**

45  **facebook.com/Brett.Bolton.333/posts/10162928980710104**

46  **facebook.com/FamilyBreakFinder/posts/3313309172016539?__tn__=-R**

did include some basic advice that is worth following, it also contained some claims which could have falsely led people to believe they have not been infected when they have, or which suggested ways of preventing infection that will not work. The post was updated to remove most of the false claims after we published the fact check, although several copied and pasted versions of the original remained. We have also seen this post, or claims from within it, in various forms circulating on Twitter and WhatsApp.

There are circumstances where an 'Unsubstantiated' and 'More Context Needed' rating would have been a useful tool earlier in the pandemic. For example, since the disease is a new one research into it is ongoing, these ratings would have been useful when there was no consensus[47] on how long the virus could survive on surfaces.

## 5G

We highlighted the risk of 5G conspiracy theories in our 2019 report, noting the distinct lack of official guidance properly addressing some public concerns, and recommending that gap was filled. We are disappointed that it has taken arson attacks[48] for the government and public health bodies to take steps to address this.

Since then a number of claims have surfaced and spread making a non-existent connection between 5G and the outbreak of Covid-19. We started seeing the link between 5G and Covid-19 claims in the second half of January, shortly after the virus started getting significant UK media coverage. There are multiple theories, which sometimes overlap, but can also contradict each other.

The queue was valuable in giving us insight into the range of claims that were being shared online. We fact checked and rated 'False' posts[49] claiming that 5G compromised human health and weakened immune systems. One of the most common false claims we saw was secret messages about 5G and Covid-19[50] (rated 'False') in the design of the new £20 note. Another prominent 'False' theory[51] is that Covid-19 symptoms were actually "mass injury" from 5G. Facebook groups, many of which existed and shared conspiracy theories about 5G before Covid-19, have been key in spreading such claims.

---

47   **fullfact.org/online/coronavirus-surfaces-packages-china**

48   **bbc.co.uk/news/technology-52281315**

49   **fullfact.org/online/wuhan-5g-coronavirus**

50   **fullfact.org/online/5g-coronavirus-20-note**

51   **facebook.com/joshyjones89/posts/10157798506676224**

Claims about 5G made up a relatively small amount of our work before the pandemic, but now contribute to a significant amount of what we see on social media. In a few weeks, we saw 5G posts go from a niche corner of the internet to several fully fledged conspiracy theories fleshed out around the world's biggest news story. This resulted in Full Fact publishing an explainer in April 2020 reviewing the background to the conspiracy theories and explaining why they were incorrect.[52]

## Understanding the impact of Third-Party Fact-Checking

We are clear that the Third-Party Fact-Checking programme is worthwhile, and we have recommended that a similar programme is introduced by other internet companies.[53]

But greater information is needed on the impact of fact checks on the prevalence of bad information and user behaviour. For example, it is not clear whether, or to what extent, Third-Party Fact-Checking is pushing users to use Facebook Messenger or closed peer to peer networks to share information instead.

In October 2020, Facebook offered fact checking partners activity summaries, which tell us how many posts with our labels were viewed each day. But without more specific data on what labels influenced what behaviour, the impact this makes is limited.

Machine readable data will help fact checkers to understand how different groups react to seeing fact checks, the decisions people take and whether this changes according to the rating given, and how behaviours might change in the long term. Anonymised data about the impact of individual fact checks would help fact checkers start to understand more about the effectiveness of our headlines, images, summaries, and to learn from long term patterns.

This should be provided at regular agreed intervals, ideally quarterly.

This additional information would help fact checkers to discuss impact and share evidence with potential funders and supporters, and allocate staff time effectively between Facebook work and other potentially impactful work such as training mainstream media journalists.

We would welcome and be prepared to contribute to more work in this area identifying the information needed from all internet companies to properly and proportionately address harmful false information online.

---

52   **fullfact.org/online/5g-and-coronavirus-conspiracy-theories-came**

53   **fullfact.org/blog/2020/apr/full-fact-report-2020**

In addition, Facebook and the other internet companies have the best information on what potential harmful misinformation is being shared on their platforms. Covid-19 misinformation has highlighted the importance of acting early to tackle harmful claims and provide accurate information. By regularly sharing insights into emerging potentially harmful misinformation trends with governments and other relevant official bodies this would improve understanding of emerging threats that could cause real world harm.

## Claim Matching

Finding repetitions of claims within text online is a process we refer to as claim matching, and it is vital to mapping the spread of misinformation online. Claim matching is necessary for large scale interventions against misinformation, but it comes with many warnings.

As a technology solution, it faces challenges in every direction. Finding repeats of an exact phrase is broadly a solved problem using existing search and document retrieval techniques, but as we broaden out to paraphrases of the claim it gets harder. Even exact matches need to be sensitive to context; the difference in someone referring to something in the positive (confirming the misinformation) or negative (disputing it) is huge. For paraphrases the complexity is considerably greater.

Some areas of repetition of misinformation can broadly be solved by existing technology. The ability of a perceptual hash to identify if an image is exactly the same as another is proven and can work at a fine grain with high certainty. Similar hashing and vector models for sentences of text are not so well developed and can not be treated simply as something that a black box AI model can compute.

As this kind of technology develops and actions are taken off the back of it, transparency will become ever more important. Different phrasing of the same claim will require different actions. As it stands there is very little public information about actions. As we move to models where repetitions and paraphrases are being found at scale, and actions are automatically taken at scale without human intervention, it is especially important that the data, definitions, choices, and their unintended consequences are carefully scrutinised.

We continue to call on all internet companies to support the creation of open regularly updated evaluation datasets to help ensure that these technologies can be responsibly applied.

## Editorial independence

Facebook has not sought to influence Full Fact's editorial choices. Facebook has never asked Full Fact to give or change any rating, or to treat any publisher in one way or another.

Facebook has asked Full Fact to fact check a specific post, a false claim that the City of London Police and Martin Lewis (a personal finance expert) are warning about a scam. Both sides were clear that the decision as to whether to do any specific fact check is Full Fact's. We had previously fact checked a variant of the same claim, and we did choose to publish a fact check on this.[54]

This notice will appear in all future reports unless there is any reason to modify it.

Facebook provides us with a queue of publicly-shared posts that Facebook has identified as potentially needing fact checking using its own systems. We do not know except in the broadest terms how these posts are chosen.

---

54  **fullfact.org/online/martin-lewis-city-of-london-police**. The previous variant did not mention Martin Lewis: **fullfact.org/online/city-of-london-police-scam**.

# Recommendations

We welcome the progress that Facebook has made in developing the programme.

In our first report, we made a number of recommendations as to the ratings system. Broadly, we suggested the creation of 'More context needed', 'Unsubstantiated', and 'Mixture' ratings that didn't reduce reach and a rating for humorous posts, outside of satire and pranks.

After just over a year, Facebook has enacted some of these recommendations, and have added a 'Missing Context' rating, added 'Altered', and replaced 'Mixture' with 'Partly false' content. However there are several areas where additional changes would make fact checkers' work more effective.

## Provide fact checkers with greater information

- In the queue, provide data points of number of shares over time that can be displayed on a graph within the product to help fact checkers understand how content is going viral.

- Invest in better claim matching to reduce the amount of false positives in the queue, and better match content to similar claims.

- Continue to collaborate with fact checkers to understand the most effective way to use the Third-Party Fact-Checking programme to prevent misinformation from spreading, particularly when considering new changes to the programme or the tools available.

## Provide publishers with more information

- Make greater information available to people when their content is fact checked, including which fact checkers have rated which post, give them the contact details of that fact checker automatically and make it mandatory for people to include a link or screenshot of the post they are referring to when making an appeal.

## Use the Third-Party Fact-Checking database

- Provide machine readable data to fact checking partners on the impact of individual fact checks, to enable partners to understand

more about the effectiveness of the programme, share evidence with funders and know how to best allocate resources.

- Share a register of emerging potentially harmful misinformation trends with governments and other relevant official bodies to help them understand emerging threats that could cause real world harm.

## Transparency and accountability

- Implement greater transparency around the use of AI in claim matching. While Facebook has provided some detail on its machine learning work this cannot be done in isolation. Substantial effort is needed by all internet companies to provide transparency on the data that powers algorithmic decision making and its effects, intended or otherwise.

**Full Fact**
**2 Carlton Gardens**
**London**
**SW1Y 5AA**

✉ team@fullfact.org

📞 +44 (0)20 3397 5140

🐦 @FullFact

🌐 fullfact.org

**FULL**
**FACT**