Facebook report on the implementation of the Code of Practice for Disinformation

Annual Report | September 2019

1. Introduction

Fighting false news and misinformation on Facebook is one of the most important things we do. Since 2016, we’ve made substantial improvements and we are constantly looking for new ways to keep false news and other types of misinformation off our platform. In the past we did not do enough to anticipate some of the risks we’ve faced, such as misinformation and election interference, so we have now made fundamental changes.

In addition to the measures we have taken to make progress under the commitments in the Code, we have also taken a number of additional steps through product and policy interventions. This past year we’ve invested record amounts in keeping people safe and strengthening our defenses against abuse. Our tactics include blocking and removing fake accounts, finding and removing bad actors, limiting the spread of false news and misinformation, and bringing unprecedented transparency to political advertising.

We’ve also more than doubled the number of independent fact-checking partners we work with, provided people with far more control over their information and the content they see in their News Feed, as well as more transparency into our policies and operations. As part of this commitment to transparency, we now release our Community Standards Enforcement Report every six months to show how we are performing against our own Community Standards along with a supplemental independent verification of our methodology. To this end, last year we established Data Transparency Group (DTAG) to provide an independent and public assessment of whether the metrics we share through CSER are meaningful and accurate.

On the technology side, we have improved our machine learning capabilities, which allows us to be more effective in finding and removing violating behavior. As a result, we’re making real progress. We remove millions of fake accounts a day, stopping them from ever engaging in the type of coordinated information campaigns often used to interfere with elections. And we’ve improved rapid response efforts across our teams. Meanwhile, we’re working far more closely with governments, outside experts and other companies.

As we continue to look at how we can improve our approach and the systems we’ve built, we reiterate our belief in freedom of expression and in the importance of people’s ability to share diverse views, experiences, ideas and information. At the same time, we also reaffirm our belief that the answer to fighting viral misinformation isn’t simply to take it down, but to prevent it from going viral. And when people do see misinformation, to help them understand the context to make informed decisions. There’s more to do. Security is an arms race. But we’re committed to the fight.

The following sections set out our activities to each category of commitments set out in the EU Code of Practice on Disinformation over the past year, as well as additional
updates on our policies, products and partnerships we have undertaken to combat misinformation. The policies, products and services detailed in the report are available globally unless specifically detailed otherwise.

2. Scrutiny of Ads Placement

Before ads appear on Facebook, they are reviewed to make sure that they meet our Advertising Policies\(^1\). During the ad review process, we check ad images, text, targeting and positioning, in addition to the content on the ads landing page before approval. Ads may not be approved if the landing page content isn't fully functional, doesn't match the product/service promoted in the advert or doesn't fully comply with our advertising policies.

Our advertising policies prohibits an extensive range of content, including but not limited to:

- Ads that violate our Community Standards\(^2\)
- Ads that constitute, facilitate or promote illegal products, services or activities\(^3\)
- Ads that discriminate against people\(^4\) based on personal attributes
- Ads that contain sensational content\(^5\), low quality or disruptive content\(^6\), misleading or false content\(^7\) and controversial content\(^8\)
- Ads that use tactics to circumvent our ad review process or other enforcement systems\(^9\)

2.1 Political and issue-based advertising

We believe that transparency leads to increased accountability and responsibility over time – not just for Facebook but advertisers as well. It’s why we continue to introduce tools that allow elected officials, those seeking office, and organizations aiming to influence public opinion to show more information about the ads they run and who’s behind them.

We have rolled out our transparency tools globally for advertisers wanting to place ads about social issues, elections or politics. (For a full list of countries and territories where these tools are now available, visit our Help Center\(^10\).)

How we detect and review political ads vary by country due to different understandings, criteria and definitions of what a political ad is. In the EU, we launched the ads transparency tools late March and define political ads as:

\(^1\) https://www.facebook.com/policies/ads
\(^2\) https://www.facebook.com/policies/ads/prohibited_content/community_standards
\(^3\) https://www.facebook.com/policies/ads/prohibited_content/illegal_products_or_services
\(^4\) https://www.facebook.com/policies/ads/prohibited_content/discriminatory_practices
\(^5\) https://www.facebook.com/policies/ads/prohibited_content/sensational_content
\(^6\) https://www.facebook.com/policies/ads/prohibited_content/low_quality_or_disruptive_experiences
\(^7\) https://www.facebook.com/policies/ads/prohibited_content/misleading_or_false_content
\(^8\) https://www.facebook.com/policies/ads/prohibited_content/controversial_content
\(^9\) https://www.facebook.com/policies/ads/prohibited_content/circumventing_systems
\(^10\) https://www.facebook.com/business/help/167836590566506
• Ads made by, on behalf of or about a current or former candidate for public office, a political party, a political action committee or advocates for the outcome of an election to public office

• Ads about any election, referendum or ballot initiative, including "get out the vote" or election information campaigns

• Ads about social issues: 1) civil and social rights; 2) economy; 3) environmental politics; 4) immigration; 5) political values and governance; and 6) security and foreign policy.

• Ads regulated by law as political advertising.

These tools are a critical part of our election integrity efforts and include:

• **Identity Confirmation and Authorization.** To help prevent abuse and foreign interference, the admins of Pages must provide an identification document\(^{11}\) that can be verified/confirmed. Admins may only be authorised in one country and must provide an ID document for the country which they would like to target for advertising.

• **Labeling Political and Issue Ads.** To increase advertiser transparency, political ads on Facebook and Instagram must be clearly labeled with a “Paid for by” disclaimer\(^{12}\). The disclaimer has several components that is verified before it is approved - page or organisation name; website; email address; phone number; and address. This information is included in each ad and publicly available on the ad.

• **Ad Library.** The Ad Library\(^{13}\) provides advertising transparency by offering a comprehensive, searchable collection of all ads that are active and running on Facebook and Instagram. This includes non-political ads. Political ads, however, are archived and remain in the Ad Library for seven years. The Ad Library shows ranges of range of impressions, range of spend, range of age and gender reached for each ad. People have the ability to report ads that they believe should have a “Paid for by” disclaimer and don’t, or for violating other Facebook advertising policies, directly from the Library. Anyone can explore the Library, with or without a Facebook account.

• **Ad Library API.** The Ad Library application programming interface (API)\(^{14}\) allows regulators, journalists, watchdog groups and other people to analyse political ads and help hold advertisers and Facebook accountable. Users can perform customized keyword searches of ads stored in the Ad Library. The results from the API include the creatives and ad performance data that is archived in the Ad Library. Since the EU launch in March, we’ve made improvements to our API so people can easily access ads from a given country and analyze specific

\(^{11}\) [http://facebook.com/id](http://facebook.com/id)


\(^{13}\) [https://www.facebook.com/business/m/one-sheeters/ad-library](https://www.facebook.com/business/m/one-sheeters/ad-library)

\(^{14}\) [https://www.facebook.com/ads/library/api](https://www.facebook.com/ads/library/api)
advertisers. We’re also working on making it easier to programmatically access ad images, videos and recently served ads.

- **Ad Library Report.** The [Ad Library Report](https://www.facebook.com/ads/library/report) is a report of aggregated insights about ads about social issues, elections or politics in the Ad Library. It provides cumulative statistics about ads in the Ad Library. For example, the report shows the total number of ads and total spend on ads in the Ad Library by country, as well as total spend by advertiser, advertiser spend by day and top searched keywords from the past week. Users can explore, filter and download the data into a CSV file. The Ad Library Report is useful to people who are interested in understanding high-level activity in the Ad Archive since launch, while the Ad Library could be used to deep dive into specific ads.

Since we first rolled out the ads transparency tools last May in the US, we are continually taking feedback, learning and improving to make them more useful. For example, we recently updated the Ad Library, Ad Library Report and Ad Library API to help journalists, lawmakers, researchers and others learn more about the ads they see. The update includes narrowing the ranges of impressions and spend in the ad details of the Ad Library; adding useful API filters; providing programmatic access to download ad creatives; and a repository of frequently used API scripts.

In November, we will begin testing a new ads database with researchers that will enable them to quickly download the entire Ad Library, pull daily snapshots and track day-to-day changes.

### 2.2 Ad Library Report Stats & Figures

From late March 2019 (when we launched the ads transparency tools for the EU27) to September 9, there were around 444,000 political ads in the EU27, totaling around 31.5 million euros of political ads spend. The following table provides a country-by-country breakdown for the EU27 taken from the Ad Library Report dated 9 September 2019:

<table>
<thead>
<tr>
<th>Ad Library Totals</th>
<th>Total Number of Political Ads</th>
<th>Total Amount Spent on Political Ads (EUR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria*</td>
<td>31,586</td>
<td>2,071,030</td>
</tr>
<tr>
<td>Belgium*</td>
<td>19,247</td>
<td>3,017,184</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>3,642</td>
<td>161,455</td>
</tr>
<tr>
<td>Croatia*</td>
<td>4,004</td>
<td>138,647</td>
</tr>
</tbody>
</table>

Includes ads related to politics or issues of importance from March 2019 to 9 Sep 2019

---

<table>
<thead>
<tr>
<th>Country</th>
<th>Population</th>
<th>GDP (Millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyprus</td>
<td>2,553</td>
<td>78,349</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>10,533</td>
<td>349,395</td>
</tr>
<tr>
<td>Denmark*</td>
<td>22,175</td>
<td>2,005,860</td>
</tr>
<tr>
<td>Estonia</td>
<td>2,500</td>
<td>99,833</td>
</tr>
<tr>
<td>Finland*</td>
<td>8,247</td>
<td>544,621</td>
</tr>
<tr>
<td>France</td>
<td>14,568</td>
<td>2,785,625</td>
</tr>
<tr>
<td>Germany</td>
<td>102,107</td>
<td>5,263,186</td>
</tr>
<tr>
<td>Greece*</td>
<td>22,014</td>
<td>1,349,986</td>
</tr>
<tr>
<td>Hungary</td>
<td>13,763</td>
<td>787,762</td>
</tr>
<tr>
<td>Ireland</td>
<td>6,581</td>
<td>441,664</td>
</tr>
<tr>
<td>Italy</td>
<td>41,173</td>
<td>3,083,968</td>
</tr>
<tr>
<td>Latvia</td>
<td>2,638</td>
<td>58,512</td>
</tr>
<tr>
<td>Lithuania*</td>
<td>3,648</td>
<td>118,527</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>3,537</td>
<td>82,440</td>
</tr>
<tr>
<td>Malta</td>
<td>4,504</td>
<td>127,414</td>
</tr>
<tr>
<td>Netherlands</td>
<td>10,888</td>
<td>961,056</td>
</tr>
<tr>
<td>Poland*</td>
<td>18,742</td>
<td>739,540</td>
</tr>
<tr>
<td>Portugal*</td>
<td>4,396</td>
<td>181,636</td>
</tr>
<tr>
<td>Romania*</td>
<td>14,898</td>
<td>856,362</td>
</tr>
<tr>
<td>Slovakia*</td>
<td>15,621</td>
<td>610,475</td>
</tr>
<tr>
<td>Slovenia</td>
<td>3,093</td>
<td>65,433</td>
</tr>
<tr>
<td>Country</td>
<td>NFC</td>
<td>HVC</td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Spain*</td>
<td>37,323</td>
<td>3,314,919</td>
</tr>
<tr>
<td>Sweden</td>
<td>19,853</td>
<td>2,188,360</td>
</tr>
</tbody>
</table>

*Countries that held or will be holding national elections in 2019, since the launch of the Ad Library in late March. Source: Ad Library Report

### 3. Integrity of Services

Authenticity is the cornerstone of our community and key to preserving the integrity of our services. We remove content and accounts that violate our Community Standards, which include explicit authenticity requirements, such as our Name Policies, and prohibitions on Misrepresentation. These policies are intended to create a safe environment where people can trust and hold one another accountable. Key aspects of these policies include prohibitions on:

- Maintaining multiple accounts
- Creating inauthentic profiles
- Sharing an account with any other person
- Creating another account after being banned from the site
- Evading the registration requirements outlined in our Terms of Service
- Creating a profile assuming the persona of or speaking for another person or entity
- Creating a Page assuming to be or speak for another person or entity for whom the user is not authorized to do so
- Engaging in inauthentic behavior, which includes creating, managing, or otherwise perpetuating:
  - Accounts that are fake
  - Accounts that have fake names
  - Accounts that participate in, or claim to engage in, coordinated inauthentic behavior, meaning that multiple accounts are working together to do any of the following:
    - Mislead people in an attempt to encourage shares, likes, or clicks

---

16 [https://www.facebook.com/communitystandards](https://www.facebook.com/communitystandards)
18 [https://www.facebook.com/communitystandards/misrepresentation](https://www.facebook.com/communitystandards/misrepresentation)
19 [https://www.facebook.com/communitystandards/integrity_authenticity](https://www.facebook.com/communitystandards/integrity_authenticity)
- Mislead people to conceal or enable the violation of other policies under the Community Standards

### 3.1 Community Standards Enforcement Report

To track our progress and demonstrate our continued commitment to make Facebook safe and inclusive, we release a [Community Standards Enforcement Report](https://transparency.facebook.com/community-standards-enforcement) every six months. The last report published was May 2019, which included Q4 2018 and Q1 2019. Updated CSER metrics are expected to be available in November 2019 for Q2 and Q3 2019.

This report shares metrics on how Facebook is performing on preventing and removing content that goes against our Community Standards. It also provides visibility into how we enforce policies, respond to data requests and protect intellectual property, while monitoring dynamics that limit access to Facebook products.

We’re continuously working on adding more metrics to the CSER. The third edition of the CSER, published May 2019, has been expanded to now include metrics for nine categories of violations: adult nudity and sexual activity, bullying and harassment, child nudity and sexual exploitation of children, fake accounts, hate speech, regulated goods, spam, terrorist propaganda, and violence and graphic content.

We have also updated the report to include how much content people appealed and how much content was restored after we initially took action.

<table>
<thead>
<tr>
<th>Policy Area</th>
<th>Amount of Content Appealed</th>
<th>Amount of Content Restored After Appeal</th>
<th>Amount of Content Restored Without Appeal</th>
<th>Total Amount of Content Restored * (Amount of Content Restored After Appeal + Amount of Content Restored Without Appeal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spam</td>
<td>20.8 M</td>
<td>5.7 M</td>
<td>38.6 M</td>
<td>44.2 M</td>
</tr>
<tr>
<td>Adult Nudity and Sexual Activity</td>
<td>2.1 M</td>
<td>453 K</td>
<td>668 K</td>
<td>1.1 M</td>
</tr>
<tr>
<td>Hate Speech</td>
<td>1.1 M</td>
<td>130 K</td>
<td>21.2 K</td>
<td>152 K</td>
</tr>
<tr>
<td>Bullying and Harassment</td>
<td>496 K</td>
<td>80.2 K</td>
<td>3.5 K</td>
<td>83.7 K</td>
</tr>
<tr>
<td>Violence and Graphic Content</td>
<td>171 K</td>
<td>23.9 K</td>
<td>45.9 K</td>
<td>69.7 K</td>
</tr>
<tr>
<td>Regulated Goods - Drugs</td>
<td>87.4 K</td>
<td>18.1 K</td>
<td>0.7 K</td>
<td>18.9 K</td>
</tr>
<tr>
<td>Regulated Goods - Firearms</td>
<td>80.8 K</td>
<td>6.2 K</td>
<td>0.7 K</td>
<td>6.9 K</td>
</tr>
<tr>
<td>Terrorist Propaganda (ISIS, Al-Qaeda and affiliates)</td>
<td>40.1 K</td>
<td>20.3 K</td>
<td>162 K</td>
<td>183 K</td>
</tr>
<tr>
<td>Child Nudity and Sexual Exploitation of Children</td>
<td>20.6 K</td>
<td>0.7 K</td>
<td>6.4 K</td>
<td>7.1 K</td>
</tr>
</tbody>
</table>

*Source: Facebook's Community Standards Enforcement Report, May 2019*
Data Transparency Advisory Group

We know it’s important to have independent verification of our methodology and work. In November last year, we established the Data Transparency Advisory Group (DTAG) — an independent body led by The Justice Collaboratory at Yale Law School comprised of international experts in measurement, statistics, criminology and governance. The purpose of this group is to provide an independent, public assessment of whether the metrics we share in the CSER, such as for fake accounts, are meaningful and accurate. We provided the advisory group detailed and confidential information about our enforcement processes and measurement methodologies. In May this year, they published their first report. In the report, the DTAG noted that the Community Standards Enforcement Report is an important exercise in transparency and that they found our approach and methodology sound and reasonable. They also highlighted other areas where we could be more open in order to build more accountability and responsiveness to the people who use our platform. These important insights will help inform our future work.

Community Standards: Product Policy Forum

We recognize how important it is for Facebook to be a place where people feel empowered to communicate and we take our role in keeping abuse off our services seriously. That’s why we have developed a set of public Community Standards that outline what is and isn’t allowed on Facebook.

The goal of our Community Standards is to encourage expression and create a safe environment. We base our policies on input from our community and from experts and organizations outside Facebook so we can better understand different perspectives on safety and expression, as well as the impact of our policies on different communities globally. Based on this feedback, as well as changes in social norms and language, our standards evolve over time.

Every two weeks, members of our Product Policy team, who sit in offices around the world, run a meeting called the Product Policy Forum to discuss potential changes to our Community Standards, ads policies and major News Feed ranking changes. A variety of subject matter experts participate in this meeting, including members of our safety, misinformation and cybersecurity policy teams, counterterrorism specialists, Community Operations employees, product managers, public policy leads and representatives from our legal, communications and diversity teams.

For transparency and to help people better understand the policies discussed, the minutes from the Product Policy Forum are made public - once a decision is made - and can be found here.

3.1.1 Fake Accounts

Preventing fake accounts is one way to stop abuse on our platforms. We focus our enforcement against abusive accounts to both prevent harm and avoid mistakenly taking action on good accounts.

In the six months between Q4 2018 and Q1 2019, we've seen a steep increase in the creation of abusive, fake accounts on Facebook. The number of fake accounts disabled spiked up from 1.2 billion accounts in Q4 2018 to 2.19 billion in Q1 2019, largely due to increased automated attacks by bad actors who attempt to create large volumes of accounts at one time. The majority of these accounts were caught within minutes of registration, before they became a part of our monthly active user population. For the accounts that are not initially detected and disabled, we find that many of them are used in spam campaigns and are financially motivated.

We try to stop fake accounts abusing our platforms in three distinct ways. Of the accounts we remove, both at sign-up and those already on the platform, 99.8% of these are proactively detected by us before people report them to us.

- **Blocking accounts from being created:** The best way to fight fake accounts is to stop them from getting onto Facebook in the first place. We’ve built detection technology that can detect and block accounts even before they are created. Our systems look for a number of different signals that indicate if accounts are created in mass from one location. A simple example is blocking certain IP addresses altogether so that they can’t access our systems and thus can’t create accounts.

  The data we include in the CSER about fake accounts does not include unsuccessful attempts to create fake accounts that we blocked at this stage. This is because we can’t know the number of attempts to create an account we’ve blocked as, for example, we block whole IP ranges from even reaching our site. While these efforts aren’t included in the report, we can estimate that every day we prevent millions of fake accounts from ever being created using these detection systems.

- **Removing accounts when they sign-up:** Our advanced detection systems look for potential fake accounts as soon as they sign-up, by spotting signs of malicious behavior. These systems use a combination of signals such as patterns of using suspicious email addresses, suspicious actions, or other signals previously associated with other fake accounts we’ve removed. Most of the accounts we currently remove, are blocked within minutes of their creation before they can do any harm.

  We include the accounts we disable at this stage in our accounts actioned metric for fake accounts. Changes in our accounts actioned numbers are often the result of unsophisticated attacks like we saw in the last two quarters. These are really easy to spot and can totally dominate our numbers, even though they pose little risk to users. For example, a spammer may try to create 1,000,000 accounts

35 https://transparency.facebook.com/community-standards-enforcement#fake-accounts
quickly from the same IP address. Our systems will spot this and remove these fake accounts quickly. The number will be added to our reported number of accounts taken down, but the accounts were removed so soon that they were never considered active and thus could not contribute to our estimated prevalence of fake accounts amongst monthly active users or our publicly stated monthly active user number or even any ad impressions.

- **Removing accounts already on Facebook:** Some accounts may get past the above two defenses and still make it onto the platform. Often, this is because they don’t readily show signals of being fake or malicious at first, so we give them the benefit of the doubt until they exhibit signs of malicious activity. We find these accounts when our detection systems identify such behavior or if people using Facebook report them to us. We use a number of signals about how the account was created and is being used to determine whether it has a high probability of being fake and disable those that are.

The accounts we remove at this stage are also counted in our accounts actioned metric. If these accounts are active on the platform, we would also account for them in our prevalence metric. Prevalence of fake accounts measures how many active fake accounts exist amongst our monthly active users within a given time period.

Updated fake account metrics are expected to be available in November 2019 for Q2 and Q3 2019.

---

**ACCOUNTS ACTIONED**

How many fake accounts did we take action on?

We removed significantly more fake accounts in Q1 2019 and Q4 2018 than in previous quarters, due to an increase in automated, scripted attacks. The majority of these accounts were caught within minutes of registration, before they became a part of our monthly active user (MAU) population.
3.1.2 Hate Speech

One of the ways we reduce misinformation on our platforms is to remove content, such as Hate Speech[^36], that violates our Community Standards. We find that content identified as hate speech would often also be seen as misinformation.

We don’t allow hate speech on Facebook because it creates an environment of intimidation and exclusion and in some cases may promote violence. We define hate speech as violent or dehumanizing speech, statements of inferiority, or calls for exclusion or segregation based on protected characteristics. These characteristics include race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disability or disease.

In the latest CSER report[^37] for Q4 2018 and Q1 2019, we noted improvements to our technology for detecting hate speech, and we expanded our detection to new languages. As a result, we increased our proactive rate - we now detect 65% of the content we remove before users reported it, up from 24% just over a year ago when we first shared our efforts. In Q1 2019, we took down 4 million hate speech posts, compared to 3.3 million in Q4 2018, and we continue to invest in technology to expand our abilities to detect this content across different languages and regions.

Updated hate speech metrics are expected to be available in November 2019 for Q2 and Q3 2019.

[^36]: https://www.facebook.com/communitystandards/hate_speech
[^37]: https://transparency.facebook.com/community-standards-enforcement#hate-speech
3.1.3 Spam

Spam\textsuperscript{38} is a broad term to describe inauthentic content and behavior on Facebook that violates our Community Standards. It can be automated (published by bots or scripts) or coordinated (when an actor uses multiple accounts to spread deceptive content). Spammers aim to build audiences to inflate their content's distribution and reach, typically for financial gain.

Content actioned for spam has increased considerably, since we found and took action on more content that goes against our standards. Spammers are adversarial and they routinely try new tactics to avoid our detection. Recently we have invested in enhanced detection technology, including improvements to our machine learning, to more accurately detect and take action on spam violations.

In both Q4 2018 and Q1 2019\textsuperscript{39}, we acted on 1.8 billion pieces of content, compared to 737 million in Q4 2017, 836 million in Q1 2018, 957 million in Q2 2018 and 1.2 billion in Q3 2018.

Updated spam metrics are expected to be available in November 2019 for Q2 and Q3 2019.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{content-actioned.png}
\caption{How much content did we take action on?}
\end{figure}

\textsuperscript{38} \url{https://www.facebook.com/communitystandards/spam}
\textsuperscript{39} \url{https://transparency.facebook.com/community-standards-enforcement#spam}
3.2 Prohibiting Coordinated Inauthentic Behaviour

Accounts, Pages and Groups removed for violations of our Coordinated inauthentic behavior\(^{40}\) (CIB) policy is a small subset of the overall accounts and content we remove from our platforms.

We’re constantly working to detect and stop CIB activity because we don’t want our services to be used to manipulate people. When we take down Pages, Groups, accounts, or Instagram accounts we remove them for misleading behaviors, not for the content they post. These behaviors include the use of fake accounts, often in coordination with real accounts, designed to mislead people about the identity of people behind an operation or the source or origin of the content.

In the run-up to the European elections, we prioritized building partnerships with civil society organizations and sharing information with government partners about the types of operations we anticipated and key opportunities for joint action to disrupt these threats.

We are making progress rooting out this abuse, but it’s an ongoing challenge. We’re committed to continually improve to stay ahead by building better technology, hiring more people and working more closely with law enforcement, security experts and other companies. We have built an expert team of analysts with experience in law enforcement, cybersecurity, and intelligence to find, investigate, and disrupt these operations. We are also growing our technical teams to build scaled solutions to finding and preventing these behaviors.

What is Inauthentic Behavior?

When people think about information operations, they often focus on the content that is being shared. Is it hate speech? Is it a threat? Is it false? But most of the content shared by coordinated manipulation campaigns isn’t provably false, and would in fact be acceptable political discourse if shared by authentic audiences.

The real issue is that the actors behind these campaigns are using deceptive behaviors to conceal the identity of the organization behind a campaign, make the organization or its activity appear more popular or trustworthy than it is, or evade our enforcement efforts. That’s why, when we take down information operations, we are taking action based on the behavior we see on our platform — not based on who the actors behind it are or what they say.

3.2.1 Inauthentic Behavior Policy Update

We recently updated our inauthentic behavior policy to further improve our ability to counter new tactics and bad actors. Here’s how we will act against a range of inauthentic activities, whether foreign or domestic, state or non-state.

**Coordinated Inauthentic Behavior (CIB)**

As with our previous takedowns, we will continue looking for groups of accounts and Pages working together to mislead people about who they are and what they’re doing. When we find domestic, non-government campaigns in which the use of fake accounts is central to the operation, we will remove all inauthentic and authentic accounts, Pages and groups directly involved in this activity. We will share our findings as part of a regular CIB report which we’ll launch in the coming months. However, if the activity we remove is directly related to a civic event, poses imminent harm or involves a new technique or a new significant bad actor, we will share our findings at the time of enforcement.

- **Persona Non Grata:** If, in the course of a CIB investigation, we determine that a particular organization is primarily organized to conduct manipulation campaigns, we will permanently remove it from our platforms in its entirety.

**Foreign or Government Interference (FGI)**

There are two types of CIB that are particularly egregious:

- Foreign-led efforts to manipulate public debate in another country
- Operations run by a government to target its own citizens. These can be particularly concerning when they combine deceptive techniques with the real-world power of a state.

If we see any instances of CIB conducted on behalf of a government entity or by a foreign actor, we will apply the broadest enforcement measures including the removal of every on-platform property connected to the operation itself and the people and organizations behind it. We will also announce the removal of this activity at the time of enforcement.

**Inauthentic Behavior (IB)**

We routinely take action against other inauthentic behaviors, including financially-motivated activity like spam or fake engagement tactics that rely on inauthentic amplification or evading enforcement, rather than a core use of fake accounts. We’ve updated a list of these deceptive techniques in our IB policy to make sure people better understand the types of inauthentic behavior we will enforce against, even if it’s not part of a CIB campaign. We enforce against IB only based on specific protocols that are reviewed and approved through our internal process. It may include temporary restrictions, warnings, down-ranking or removal.

We will continue to adapt our policies to ensure we can effectively combat information operations even as bad actors evolve their techniques.

**CIB Network Disruptions**

In the run-up to the European elections, we disrupted the following CIB cases within Europe and its surrounding neighbourhood:

41 [https://www.facebook.com/communitystandards/inauthentic_behavior](https://www.facebook.com/communitystandards/inauthentic_behavior)
Since the election, we disrupted the following CIB cases within Europe and its surrounding neighbourhood:

- **Removing Coordinated Inauthentic Behavior in Thailand, Russia, Ukraine and Honduras** [49] (July 25)

  For Ukraine specifically, we had two separate takedowns of assets originating from Russia. A total of 149 Facebook and Instagram accounts, pages and groups were removed. We shared information about our analysis with law enforcement, policymakers and industry partners.

- **Removing Coordinated Inauthentic Behavior in UAE, Egypt and Saudi Arabia** [50] (August 1)

  We removed 259 Facebook accounts, 102 Facebook Pages, five Facebook Groups, four Facebook Events and 17 Instagram accounts that originated in the UAE and Egypt and focused on a number of countries, primarily in the Middle East and some in North and East Africa, including Libya, Sudan, Comoros, Qatar, Turkey, Lebanon, Syria, Jordan and Morocco.

We also removed 217 Facebook accounts, 144 Facebook Pages, five Facebook Groups and 31 Instagram accounts that originated from Saudi Arabia and focused primarily on the Middle East and Northern Africa, including Qatar, Saudi Arabia, UAE, Bahrain, Egypt, Morocco, Palestine, Lebanon and Jordan.

- **Removing Coordinated Inauthentic Behavior From Iraq and Ukraine** (September 16)
We removed 269 Pages, 80 Groups and 244 accounts from two separate, unconnected operations that originated in Iraq and Ukraine. We didn’t find any links between the campaigns we removed, but they both created networks of accounts to mislead others about who they were and what they were doing.

The network in Iraq was domestic-focused and typically focused about domestic political and societal issues such as religion, various public figures including Saddam Hussein, the state of the military under the Saddam rule, tensions with Iran, the US military action in Iraq, Iranian-backed militia operating in Iraq and Kurdish-Iraqi politics.

The in Ukraine typically posted about celebrities, show business, sports, local and international news, political and economic issues including Ukrainian elections, political candidates and criticism of various public figures. Although the people behind the activity in Ukraine attempted to conceal their identities, our review linked this activity to Pragmatico, a Ukrainian PR firm.

- **Removing Coordinated Inauthentic Behavior in Spain** (September 20)

  We removed 65 Facebook accounts and 35 Instagram accounts, as part of a domestic-focused network in Spain. The individuals behind this activity used fake accounts — many of which had already been disabled by our automated systems — to comment on, amplify and disseminate their own content and posts by others. While most of this activity was focused on amplifying others’ content, some of the account owners also posted about local and national political news and issues including topics like elections, the Partido Popular Party and criticism of its political opponents. Although the people behind this activity attempted to conceal their identities, our investigation found links to individuals associated with Partido Popular.

- **Removing More Coordinated Inauthentic Behavior From Iran and Russia** (October 21)

  We removed four separate networks of accounts, Pages and Groups for engaging in coordinated inauthentic behavior on Facebook and Instagram. Three of them originated in Iran and one in Russia, and they targeted a number of different regions of the world: the US, North Africa and Latin America. All of these operations created networks of accounts to mislead others about who they were and what they were doing. We have shared information about our findings with law enforcement, policymakers and industry partners.

### 3.2.2 Integrity & Security Initiative

Working more closely with law enforcement, security experts and industry partners is an important part of our strategy to root out bad actors on our platforms. To facilitate closer collaboration and partnership, we launched the Integrity & Security Initiative (ISI) together with the German Federal Office for Information Security (BSI) at the start of the year.
The aim of ISI is to create a better and more comprehensive understanding of interference into elections and develop guidance on how to combat election interference. Since the launch, Google and the Dutch Ministry of the Interior have also joined ISI. Meetings took place in February, April and June. In the last two sessions, ISI focused on the integrity of the European Elections.

The session in April focused on the exchange of views and information on potential threats for the European Elections, updates on measures taken by the platforms, technical overview on how platforms deal with information operations and identifying fake accounts, and ended with a commitment by platforms to implement additional security measures. The session in June was used to discuss observations and learnings from the European Elections as well as a program to improve cybersecurity for politicians on social media ("Politiker sicher im Netz").

3.2.3 Election Security Escalation Channel

In the run up to the European Elections, we launched an Election Security Escalation Channel for trusted security and intelligence partners to directly report potential cybersecurity and information operations threats. This reporting channel includes partners from eleven EU countries and continues to be maintained post-election. It allows us to respond more rapidly to escalations and threats.

3.3 Reducing and Removing False News and Misinformation

Our approach to misinformation is guided by the principle that we should work to provide people with informative information, while balancing free expression. However, we also know that our users want to see high quality content on our platform which is why our strategy to combat misinformation has three parts: remove, reduce, inform51.

- We remove content that violates our community standards, which enforce the safety and security of the platform.
- We reduce the distribution of content that does not directly violate our Community Standards, but still undermines the authenticity of the platform, by demoting it in News Feed.
- We inform our users and empower people to decide for themselves what to read, trust and share by giving them more context in-product.

We think that reducing the distribution of misinformation—rather than removing it outright—strikes the right balance between giving people a place to express themselves and creating a community that’s safe and authentic. There are some extreme forms of misinformation that we remove from our platform, including misinformation that leads to offline violence and misinformation that leads to voter suppression. Across these areas, we have made significant progress in developing and enforcing our policies over the past year.

3.3.1 Removing Misinformation that Contributes to Offline Violence

When it comes to misinformation that has the potential to contribute to offline violence, we have learned that in some cases, people post content that, on its face, doesn't violate. However, coupled with an understanding of ethnic and religious tensions on the ground or a more complete view of the publisher and the target, seemingly benign online content has the potential to incite or cause real world harm. To address this reality, we updated our Community Standards such that we remove misinformation that may contribute to imminent violence or physical harm. To operationalize the policy, we have onboarded local organizations and international institutions, who are often the first to become aware of inaccurate or misleading information that may contribute to real world harm. We have secure channels for these organizations to reach out to us and escalate content quickly.

3.3.2 Removing Misinformation that Contributes to Voter Suppression

As part of our ongoing efforts to prevent people from misusing Facebook during elections, we broadened our policies last fall against voter suppression - actions that are designed to deter or prevent people from voting. The policy now covers:

- Misrepresentation of the dates, locations, times and methods for voting or voter registration (e.g. “Vote by text!”);
- Misrepresentation of who can vote, qualifications for voting, whether a vote will be counted and what information and/or materials must be provided in order to vote (e.g. “If you voted in the primary, your vote in the general election won’t count.”); and
- Threats of violence relating to voting, voter registration or the outcome of an election.

We also introduced a new reporting option on Facebook so that people can let us know if they see voting information that may be incorrect, and have set up dedicated reporting channels.

---

52 https://www.facebook.com/communitystandards/credible_violence
We remove this type of content regardless of who it’s coming from. We also recognize that there are certain types of content, such as hate speech, that are equally likely to suppress voting. That’s why our hate speech policies ban efforts to exclude people from political participation on the basis of things like race, ethnicity or religion (e.g., telling people not to vote for a candidate because of the candidate’s race, or indicating that people of a certain religion should not be allowed to hold office).

We recognize that some posts that are reported to us may require additional review. For example, we’re unable to verify every claim about the conditions of polling places around the world (e.g. “Elementary School Flooded, Polling Location Closed”). In these cases, we will send content to our third-party fact-checkers for review. Content that is rated false will be ranked lower in News Feed, and accompanied by additional information written by our fact-checkers (what we call, Related Articles) on the same subject. We are also continuing to expand and develop our partnerships to provide expertise on trends in voter suppression and intimidation, as well as early detection of violating content.

3.3.3 Reducing the Distribution of False News in News Feed

As mentioned above, we think that in most cases reducing the distribution of misinformation - rather than removing it outright - strikes the right balance between encouraging free expression and promoting a safe and authentic community, and we believe that down-ranking inauthentic content strikes that balance. In other words, we allow people to post it as a form of expression, but we’re not going to show it at the top of News Feed.

When we demote stories rated false by fact checkers, we are able to reduce its future views. News Feed is ranked based on hundreds of thousands of signals, and some signals can identify clickbait, false news or engagement bait, which can factor into that post’s ranking in a negative way if identified. Other signals we use to identify false content includes factors like the content’s source, who is sharing it, and how people engage with the content.

Once a story has been debunked and its distribution reduced in News Feed, we use machine learning to help us identify duplicates of that story. For example, after a fact-checker in France debunked the claim that you can save a person having a stroke by using a needle to prick their finger and draw blood, we were able to identify over 1,400 links and over 20 domains spreading the same claim, and reduce their spread.

3.3.4 Fact-Checking

Today we have 56 partners in 45 languages, which is more than double where we were a year ago, and we’re working hard to continue expanding this program across Europe and globally. Since we signed the EU Code of Practice on Disinformation, we have expanded our program by onboarding new partners in Spain, Poland, Portugal, Greece, Northern Ireland, Croatia and Lithuania. Our program now includes 23 partners fact-checking content in 14 European languages: Croatian, Danish, Dutch, English, French, German, Greek, Italian, Lithuanian, Norwegian, Polish, Portuguese, Spanish and

Swedish. We recently added fact-checkers, Polígrafo, in Portugal and, Demagog, in Poland and will continue expanding our fact-checking network across the EU and beyond.

Our program works by identifying content to be reviewed by our fact-checkers through a combination of technology and human review. Our machine learning models identify links to articles which might be false and we can use the model predictions to prioritize the links we show third-party fact-checkers. Since there are hundreds of millions of links per week being shared on Facebook, we prioritize third-party fact-checkers’ time with algorithms to surface suspicious links. Fact-checkers can also proactively identify the content they would like to review themselves. These fact-checkers then review the stories, check their facts, and rate their accuracy. False stories are demoted in News Feed so fewer people see them. These confirmed fact-checker ratings then help further train our machine learning model, creating a continuous cycle of improvement.

In terms of improvements we’ve made to our program, we have taken a number of additional steps including:

- **Preventing potential misinformation from going viral**: One of the things we’ve been working on is preventing potential misinformation from going viral, especially given that quality reporting and fact-checking takes time. In many countries, if we have signals that a piece of content is false, we temporarily reduce its distribution pending review by a third-party fact-checker.

- **Reaffirming our policies related to politicians**: We believe that people should be able to see what politicians share on Facebook, so posts from politicians will not be subject to ratings from our third-party fact-checking partners. However, when a politician shares previously debunked content including links, videos and photos, we will demote that content, display related information from fact-checkers, and reject its inclusion in advertisements. Fact-checkers will continue to fact-check content about politicians and we will take action by reducing the reach of these false stories and informing people with additional context in-product. Any ads from politicians must still comply with our Community Standards and Advertising policies, including new standards that require registration and transparency.

### 3.3.5 Reducing Economic Incentives

When it comes to fighting false news, one of the most effective approaches is removing economic incentives for traffickers of misinformation. We’ve found that a lot of fake news is financially motivated. These spammers make money by masquerading as legitimate news publishers and posting hoaxes that get people to visit their sites, which are often mostly ads.

For Pages and domains that repeatedly share false news, their distribution will be reduced and their ability to monetize and advertise will be removed. These repeat offenders will no longer be allowed to advertise on Facebook. This policy includes Pages that are sharing false news but didn’t create it themselves because we believe that they,  

---

like publishers, have a responsibility for the content they share with their audiences. Once pages and domains are no longer repeat offenders, we no longer reduce their distribution and once again allow them to advertise and monetize.

3.3.6 Addressing Health-Related Misinformation

People come together on Facebook to talk about, advocate for, and connect around things like nutrition, fitness and health issues. But in order to help people get accurate health information and the support they need, it’s imperative that we minimize health content that is sensational or misleading.

In our ongoing efforts to improve the quality of information in News Feed, we consider ranking changes based on how they affect people, publishers and our community as a whole. We know that people don’t like posts that are sensational or spammy, especially about health issues. So we have made two ranking updates to reduce (1) posts with exaggerated or sensational health claims and (2) posts attempting to sell products or services based on health-related claims.

We are also working to tackle vaccine misinformation on Facebook by reducing its distribution and providing people with authoritative information on the topic. Leading global health organizations, such as the World Health Organization and the US Centers for Disease Control and Prevention, have publicly identified verifiable vaccine hoaxes. If these vaccine hoaxes appear on Facebook we take a series of steps:

- We will reduce the ranking of content from groups and Pages that spread misinformation about vaccinations in News Feed. These groups and Pages will not be included in recommendations or in predictions when you type into Search.

- When we find ads that include misinformation about vaccinations, we will reject them. We also removed related targeting options, like “vaccine controversies.” For ad accounts that continue to violate our policies, we may take further action, such as disabling the ad account.

- We may also remove access to our fundraising tools for Pages that spread misinformation about vaccinations on Facebook.

- We believe in providing people with additional context so they can decide whether to read, share, or engage in conversations about information they see on Facebook. We are exploring ways to give people more accurate information from expert organizations about vaccines, as shown in the below photo.

---

3.4 Protecting Facebook Live Videos

We recognize the tension between people who would prefer unfettered access to our services and the restrictions needed to keep people safe on Facebook. Our goal is to minimize risk of abuse on Live while enabling people to use Live in a positive way every day. We’ve been reviewing what more we can do to limit Facebook Live from being used to cause harm or spread hate and misinformation. As a result, people who have broken certain rules on Facebook — including our Dangerous Organizations and Individuals policy — will be restricted from using Facebook Live. We are also applying a ‘one strike’ policy to Live in connection with a broader range of offenses. Anyone who violates our most serious policies are restricted from using Live for set periods of time – for example 30 days – starting on their first offense, and we have extended these restrictions to other areas, such as preventing those same people from creating ads on Facebook.

3.5 Election Operations Center

Building on lessons we’ve learned over the past few years, we have established regional operations centers that provide 24/7 coverage, focused on election integrity, in Menlo Park, Singapore and Dublin. The aim of the Election Operations Centers is to help strengthen global and cross-functional coordination and speed up response times in the final, critical weeks leading up to an election.

Each hub is staffed by experts and leads from a broad range of teams - such as threat intelligence, data science, engineering, research, community operations, legal and other teams - and serve as an added layer of defense against false news and misinformation, hate speech, voter suppression, other Community Standards violations, and election interference.

57 https://www.facebook.com/communitystandards/dangerous_individuals.organizations
For the European Elections, over 500 full-time employees and over 40 regional and functional teams across the company were plugged into the Elections Operation Centers.

4. Empowering Consumers

We want to empower people to decide for themselves what to read, trust, and share, and we do this by providing greater transparency and control for people on Facebook, as well as promoting media literacy.

4.1 Labeling Photo and Video Misinformation

Misinformation is something that affects all of us, which is why it is critical to provide people with more context to make informed decisions. To help do this, we are investing in new ways to provide more context to users about the information they are seeing on the platform. We have done this in a variety of ways:

- **Notifying Users**: For Page owners, we notify them if they have shared something that was given a “false” or “mixture” rating by a fact-checker. We’ll also notify people who are about to share or have shared that post in the past.

- **Related Articles**: We developed a feature to give people more information about the publishers and articles they see. For example, when a fact-checker writes an article that explains why a claim is false, misleading, or true, we’ll show that explainer article alongside the content in question in News Feed.

- **Context Button**: Last year we launched a new feature called the Context Button[^58] to make it easier for people to view more information about websites and publishers they see on Facebook. This spring, we expanded the Context Button to images that have been fact-checked on Facebook.

Our users have asked us for help in recognizing misinformation and informing themselves about misinformation when they do come across it on the platform. This is why, in addition to the strategies above, we are also now evolving our existing related article units through a new design to better warn people when content has been labeled false or misleading by third-party fact-checkers.

Starting in the US and expanding to the rest of the world, people will now see a warning on photos and videos that have been fact-checked as false or misleading. The labels 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. Much like we do on Facebook when people try to share known misinformation, we’re also introducing a new pop-up that will appear when people attempt to share posts on Instagram that include content that has been debunked by third-party fact-checkers.

4.2 Greater Transparency and Control

Whether you’re new to Facebook or have been using it for years, you should be able to easily understand and adjust how your information influences the ads and posts you see. This is why we have introduced tools and continue to add more information and features so that people can have more context and control over what they see on Facebook.

4.2.1 Page Transparency

We want to make sure people are using Facebook authentically, and that they understand who is speaking to them. We believe that when people visit a Page or see an ad on Facebook, they should be able to understand more about the Page. This is why we continue to improve and add more information about Pages so that people can scrutinize and hold those managing a Page to account.

In March, we added more information about Pages, including the date the Page was created; whether the Page name has changed; whether the Page has merged with other pages; and the primary country locations of people who manage the Page. This gives people more context on the Page and makes it easier to understand who’s behind it.

4.2.2. Why Am I Seeing This?

Why am I seeing this ad? shows people reasons why they are seeing an ad on Facebook. In the past, ‘Why am I seeing this ad?’ highlighted one or two of the most relevant reasons, such as demographic information or that they may have visited a website. Now, people can see more detailed targeting, including the interests or categories that matched them with a specific ad. It will also be clearer where that

60 https://www.facebook.com/help/562973647153813?helpref=faq_content
information came from (e.g. the website they may have visited or Page they may have liked), and we’ll highlight controls people can use to easily adjust their experience.

We have also updated the Ad Preferences tool to show people more about businesses that upload lists with their information, such as an email address or phone number. People now see a tab with two sections:

- Advertisers who uploaded a list with the person’s information and advertised to it. This section includes advertisers that uploaded a list with the person’s information and used that list to run at least one ad in the past seven days. For example, a fitness studio that uploaded a list of client emails and used that for advertising could show up in this section.

- Businesses who uploaded and shared a list with people’s information. This section aims to help you understand the third parties and businesses who have uploaded and shared lists with people’s information. In this section, people will see the business that initially uploaded a list, along with any advertiser who used that list to serve the person an ad within the last 90 days.

For organic posts, we have introduced Why am I seeing this post? Similar to the ad version of this tool, it aims to help people better understand and more easily control what they see from friends, Pages and Groups in their News Feed. This is the first time that we’ve built information on how ranking works directly into the app.

The goal of News Feed is to show people the posts that are most relevant to them, so “Why am I seeing this post?” (which can be found in the drop down menu in the right hand corner of a post) explains how your past interactions impact the ranking of posts in your News Feed. We’ve also added shortcuts to controls, such as See First, Unfollow, News Feed Preferences and Privacy Shortcuts, to help people more easily personalize their News Feed.

61 https://www.facebook.com/ads/preferences
4.2.3 Off-Facebook Activity

Many apps and websites are free because they’re supported by online advertising. And to reach people who are more likely to care about what they are selling, businesses often share data about people’s interactions on their websites with ad platforms and other services. This is how much of the internet works, but given that the average person with a smartphone has more than 80 apps and uses about 40 of them every month, it can be really difficult for people to keep track of who has information about them and what it’s used for.

To help shed more light on these practices that are common yet not always well understood, we have introduced a new way for people to view and control their off-Facebook activity. **Off-Facebook Activity**[^63] lets people see and control the data that other apps and websites share with Facebook. People can:

- See a summary of the information other apps and websites have sent Facebook through our online business tools, like Facebook Pixel or Facebook Login;
- Disconnect this information from their own account if they want to; and
- Choose to disconnect future off-Facebook activity from their own account. People can do this for all of their off-Facebook activity, or just for specific apps and websites.

In Europe, we are starting by gradually making Off-Facebook Activity available to people in Ireland and Spain. We will continue to roll it out everywhere over the coming months to help ensure it’s working reliably for everyone.

4.3 Media & Digital Literacy

Part of our work to stop the spread of misinformation is helping people spot it for themselves. That’s why we partner with organizations and experts in media literacy. We want to help people make informed decisions on the information they’re reading and empower them to decide for themselves what to read, trust, and share. We do so by partnering with local organisations, sponsoring media literacy programs, and sharing tips to help spot false news[^64].

- **Pan-EU**: For the European Elections, we supported fact-checking partners across the E.U. to launch a media literacy campaign that ran in all 28 members states and provided tips on **how to stamp out false news**[^65]. We continue to run this campaign across the EU where there is an upcoming election. According to details provided by our agency partners - and not independently verified by Facebook - this campaign received over 100 million impressions across the EU, and our partners continue running this campaign where there is an upcoming election.

[^64]: https://www.facebook.com/help/188118808357379
[^65]: https://fullfact.org/toolkit/#three-questions
We also partnered with EAVI (the European Association for Viewers Interests) to participate in Yo!Fest (a youth event run by the European Youth Forum) to hold two media literacy training sessions and a media literacy pop quiz session.

And finally, we partnered with Freeformers and over 20 in-country NGOs and training organisations to deliver a Digital Skills Training Programme to 75,000 citizens across seven countries in Europe: Germany, France, Italy, Spain, Poland, the United Kingdom and Ireland. The series of workshops were designed to develop the confidence and skills people need in the digital economy, and to become more enthusiastic to learn more - including in the media literacy space to ensure they are safe online.

- **Denmark**: Early this year, we organised events for final year high-school students in Copenhagen and Aarhus to reach first-time voters in the Danish election, and educating them on how to navigate critically and safely social media and the internet. Discussions about filter-bubbles, freedom of speech and false news was a focal point at the events.

- **Germany**: We launched an inaugural event in June with the newspaper ZEIT - as part of their cooperation programme "ZEIT für die Schule“ - educating students on identifying false news and finding reliable information. The event presented awards to five schools participating in the programme, and we also distributed learning materials on news literacy to around 120,000 teachers.

- **Poland**: We worked with the Digital Youth Forum in Poland on a three-day educational event for youth that aimed to promote safe, informed and innovative use of new technologies as an alternative to risky online behaviours. 400 youths, between ages 13-17, participated in the Forum and over 10,000 students from 148 schools all over Poland also followed the event online. “Przestrzeń” from Facebook, the Warsaw community hub, also held digital skills and safety workshops for 94 teens as part of this initiative.

We recently launched a “Fixing the Internet” initiative with PISMO, which is part of a media literacy education program. The event gathered over 120 representatives from Warsaw-based media, NGOs and policy stakeholders, as well as a select group of Pismo magazine readers.

- **Sweden**: Facebook hosted a delegation of public and private representatives from all levels of the educational sector in Sweden - including school authorities, the Teachers' Union, companies, MPs and the Research Institute for Innovation Sweden - at its headquarters in California to discuss media literacy and disinformation. The session focused on sharing best practices and discussing common efforts that could help increase people’s resistance against disinformation, propaganda and online hateful content.

- **Digital Literacy Library**: Last autumn, we launched the Digital Literacy Library (DLL) in 45 languages around the world. The DLL is a collection of ready-to-use lessons from the Youth and Media team at the Berkman Klein Center for

---

66 [https://www.facebook.com/safety/educators](https://www.facebook.com/safety/educators)
Internet & Society at Harvard University, aimed to help young people between the ages of 11-18 to think critically and share thoughtfully online. The lessons address thematic areas such as privacy and reputation, identity exploration, security, safety and wellbeing, and more. This library is a resource for educators looking to address digital literacy and help these young people build the skills they need to safely enjoy digital technology.

- **Facebook Youth Portal**: Last May, we launched our [Youth Portal](https://www.facebook.com/safety/youth), a central place for teens that includes:
  - **Education**: Information on how to get the most out of products like Pages, Groups, Events, and Profile, while staying safe. Plus, information on the types of data Facebook collects and how we use it
  - **Peer Voices**: First person accounts from teens around the world about how they are using technology in new and creative ways
  - **Ways to control your experience**: Tips on things like security, reporting content, and deciding who can see what you share
  - **Advice**: What to do if you need a social media break, and some guidelines for how to get the most out of the internet

The portal is now available in 60 languages at [facebook.com/safety/youth](https://facebook.com/safety/youth) and we’re exploring new ways to bring these tips directly to teens on Facebook. We’re also talking to teens around the world to learn from them about how they use Facebook and working with policymakers, privacy experts, other companies and user experience designers to give teens the tools and information they need.

### 4.4 Labeling State-Controlled Media

We want to help people better understand the sources of news content they see on Facebook so they can make informed decisions about what they’re reading. Starting in November 2019, we’ll begin labeling media outlets that are wholly or partially under the *editorial control* of their government as state-controlled media. This label will be on both their Page and in our Ad Library.

We will hold these Pages to a higher standard of transparency because they combine the opinion-making influence of a media organization with the strategic backing of a state.

We developed our own definition and standards for state-controlled media organizations with input from more than 40 experts around the world specializing in media, governance, human rights and development. Those consulted represent leading academic institutions, nonprofits and international organizations in this field, including Reporters Without Borders, Center for International Media Assistance, European Journalism Center, Oxford Internet Institute’s Project on Computational Propaganda, and [other experts](https://newsroom.fb.com/news/2019/10/update-on-election-integrity-efforts/).
Center for Media, Data and Society (CMDS) at the Central European University, the Council of Europe, UNESCO and others.

It’s important to note that our policy draws an intentional distinction between state-controlled media and public media, which we define as any entity that is publicly financed, retains a public service mission and can demonstrate its independent editorial control. At this time, we’re focusing our labeling efforts only on state-controlled media.

We will update the list of state-controlled media on a rolling basis beginning in November. And, in early 2020, we plan to expand our labeling to specific posts and apply these labels on Instagram as well. For any organization that believes we have applied the label in error, there will be an appeals process.

4.5 Training and Stakeholder Support

In the run up to the elections, we provided safety and security, best practice and standard product training to electoral stakeholders to educate them on safety and security features on Facebook, reporting and escalation mechanisms, best practice for Page admins, and our election integrity work.

**Newsrooms & Journalists Training:** To help newsrooms and journalists prepare for the European Elections, our News Partnership team conducted 10 training events across the EU from March through May. We trained over 400 journalists on how to tell digital stories on our platforms, how to spot false news and how to maintain integrity in digital reporting. In September, we hosted a Publisher’s Day in Portugal. A total of 150 participants attended a 3-day event, where we talked about integrity and safety for journalists and provided training on Crowdtangle and other products for news publishers and journalists.

**Reporting Channel for Political Pages Admins:** We launched a dedicated escalation channel for admins of government and political Pages. Called the Government, Political and Advocacy Concierge (GPAC), the escalation channel allows political Pages admins to report tech issues, bugs, imposter accounts, violating and suspicious content, and other issues.

**Best Practice, Safety & Security:** To ensure electoral stakeholders - such as political parties, political groups, elected representatives and their staff, candidates, government institutions and civic organisations - are informed on safety and security best practices as well as our election work, we provided numerous training and help desk sessions across the EU. We also ran information units on people’s News Feeds about tools and resources to help political parties, politicians and candidates stay secure on our platform and established a website - [facebook.com/gpa](http://facebook.com/gpa) - with information specifically for government and political stakeholders.

4.6 Supporting an Informed Electorate

Supporting an informed electorate is an important part of our civic integrity work. People are already using Facebook to talk about politics and issues that matter to them and to communicate with their elected officials. We want to support them and make it easier to connect them to reliable information about elections. We do so by rolling out News
Feed units, such as the Election Day Reminder to direct people to official information (e.g. about polling locations). The following are the news feed units we ran for the European Elections:

- **Facebook Voter Information Reminder:** In the days leading up to the election, we ran a Voter Information Reminder across all 28 EU countries. The reminder was displayed at the top of people’s News Feed and included a button to see more information about the upcoming election in their country. The button would redirect people to either the [EU election website](http://eu-elections.eu) or the official election website of the country.

- **Election Day Reminder:** On election day, we ran an Election Day Reminder in each member state, reminding people that it is voting day. The reminder was displayed at the top of people’s News Feed and included a button to see more information, which would redirect people to either the [EU election website](http://eu-elections.eu) or the official election website of the country. The reminder also included a “Share That You’ve Voted” button, prompting people to engage and spread the word about voting.

- **Instagram Story EU Elections Sticker:** We created two unique Instagram Story stickers for the upcoming European parliamentary election, including one that says “I voted” in 24 different languages. We worked with the European Parliament on the design of the stickers. The stickers were made available in all 28 EU countries during the week of elections, from May 22 to 26. The sticker allowed Instagram users to share the historic moment with their friends and followers. The stickers included a reminder that people could tap and be directed to the [eu-european-elections.eu](http://eu-european-elections.eu) website for more information on how to vote.

5. **Empowering the Research Community**

Facebook invests in research as part of our overall efforts to make the internet and people on Facebook more secure. It helps us develop a foundational understanding of how best to serve our community by building better products and offering valuable services. In addition, the work of independent researchers is contributing to our understanding of trends in the misinformation space in particular.

Our largest effort in such research, for example - [Social Science One](https://socialscienceone.org) (SS1) - aims to deepen public understanding of the role social media has on elections and democracy and help Facebook and other companies improve their products and practices. Over the past two years, we have made significant improvements in how we monitor for and take action against abuse on our platform. We know we can’t do this work alone, and much of the progress we have made is due to significant support from external partners, including governments, civil society groups, NGOs, other private sector companies and academics. SS1 will deepen our work with universities around the world as we continue to improve our ability to address current threats and anticipate new ones.

---

69 [https://socialscience.one/](https://socialscience.one/)
5.1 Social Science One

In April 2018, we launched a partnership between Facebook, Social Science One (SS1) - a commission of 83 academic researchers throughout the world - and a group of funders, with the goal of building the right processes to share data with the academic research community, while maintaining stringent privacy protections for people who use our services. Since our initial launch, Facebook has employed nearly 20 people full-time across the company to support this effort.

To date, more than 60 researchers from 30 academic institutions across 11 countries were chosen through a competitive peer review process organized by the Social Science Research Council (SSRC). The full list of research grants awarded and other details can be found on the SS1 website. To assure the independence of the research and the researchers, Facebook does not play any role in the selection of the individuals or their projects and will have no role in directing the findings or conclusions of the research.

Additionally, in support of this effort, Facebook has built a first-of-its-kind data sharing infrastructure to provide researchers access to Facebook data in a secure manner that protects people’s privacy. We’ve consulted with some of the world’s leading external privacy advisors and the SS1 privacy committee for recommendations on how best to ensure the privacy of the data sets shared and have rigorously tested our infrastructure to make sure it is secure. Some of these steps include building a process to remove personally identifiable information from the dataset, only allowing researcher access to the dataset through a secure portal that leverages two-factor authentication and VPN.

In addition to building a custom infrastructure, Facebook has also been testing the application of differential privacy, which adds statistical noise to raw datasets to ensure an individual can’t be re-identified while simultaneously allowing for meaningful aggregate results. It also ensures the system cannot be repeatedly queried to circumvent privacy measures. We hope that this testing will help us offer more datasets to the research community.

We understand many stakeholders are eager for data to be made available as quickly as possible. We agree - while we remain committed to advancing this important initiative, we are also committed to taking the time necessary to incorporate the highest privacy protections and building a data infrastructure that provides data in a secure manner. With these safeguards in place, selected researchers received access to the following dataset in June 2019:

- **CrowdTangle**: CrowdTangle allows researchers to track the popularity of news items and other public posts across social media platforms. The CrowdTangle API will allow researchers to access public Facebook and Instagram data, which includes posts from public pages, public groups and verified profiles.

- **Ad Library API**: The Ad Library API provides data on ads related to politics or issues on Facebook in the US, UK, Brazil, India, Ukraine, Israel and the EU.

---

71 https://socialscience.one
72 https://research.fb.com/blog/2019/06/preserving-privacy-while-fostering-meaningful-research-on-elections-and-democracy/
Researchers also have access to the API. In addition, Facebook and SS1 are also working to collect feedback on the API from researchers around the world to help make it more useful for research purposes.

- **Facebook URLs Data Set:** The URL dataset is aggregated and anonymized to prevent researchers from identifying any individual Facebook users. This data set includes URLs that have been shared on Facebook by at least 100 unique Facebook users who have posted the URL with public privacy settings. This dataset includes the URL and information on the total shares for a given URL, a text summary of content within the URL, engagement statistics such as the top country where the URL was shared, and information related to the fact-checking ratings from our third-party fact-checking partners. More details on what is contained in this data set can be found in the [URL Codebook](https://socialscience.one/url-shares-codebook).

Researchers have already started making use of the data we provided and have published reports. For example, with CrowdTangle data, the University of Urbino Carlo Bo in Italy have recently published a report on coordinated and inauthentic link sharing behavior in Italy’s 2018 General Election and 2019 EU Election.

In addition to the datasets and infrastructure that we have released to the academic community, we have also had significant accomplishments in building the processes necessary to scale these data sharing efforts with the goal of providing even more rigorous datasets to support the academic community.

Perhaps most importantly, we created and negotiated a [Research Data Agreement](https://socialscience.one/research-data-agreement) to protect our users, the researchers, and their universities. This agreement has now been signed by approximately 20 universities, which sets the stage for future data sharing and collaborative research on social media data. Through these agreements, we are now sharing data about public content on Facebook and about users’ engagement with that content. We are currently sharing posts shared on public pages, groups, and verified profiles. This dataset contains public pages, and groups that are associated with political parties, politicians, elections, or causes of current controversy.

In addition, we are currently sharing all engagement statistics including clicks, likes, shares for public posts shared on public pages, groups and verified profiles. In addition, some reach data is also available to researchers via CrowdTangle, including the number of followers of a page, from the time it was first tracked onwards.

While this progress is significant and useful to advance public understanding, we also recognize that we have to do much more to realize the ambitious vision set last year. Over the coming months, we plan to provide additional data. We believe this data will allow researchers to answer even more robust questions related to exposure to disinformation across our platforms.

---

73 [https://socialscience.one/url-shares-codebook](https://socialscience.one/url-shares-codebook)

74 [https://osf.io/preprints/socarxiv/3jteh](https://osf.io/preprints/socarxiv/3jteh)

75 [https://socialscience.one/research-data-agreement](https://socialscience.one/research-data-agreement)

76 The RDA is the foundational agreement we are using to work with independent academic researchers, and is a part of our Social Science One work. It is not exclusively for Social Science One researchers – it can be used with any academic researcher who wishes to gain access to Facebook data.
While we understand that everyone, including Facebook, wishes this process could move faster, we are working as fast as is responsibly possible, and are entirely committed to this work. That is why Facebook has repeatedly and publicly committed to continue the work on data access for researchers.

For more information about the SS1 Facebook Project, go to https://socialscience.one/faq-fb.

5.2 Research Grants & Awards

Every year, we invest in numerous research projects as part of our overall efforts to make the internet and people on Facebook more secure. We also look for research that will help develop a foundational understanding of how best to serve our community and contribute to our understanding of trends in the misinformation space. The following are some of the research, aside from SS1, we have supported for this purpose. For more information on our research initiatives, go to research.fb.com.

Facebook Fellowship Program: Launched in 2010, our Fellowship Program aims to foster ties to the academic community and support the research of promising doctoral students who are engaged in innovative and relevant research in areas related to computer science and engineering at an accredited university. Fellows of the program are entitled to receive two years of tuition and fees paid, a stipend of $37,000 each year and up to $5,000 in conference travel support. Every year, we organise a Fellowship Summit at our headquarters in Menlo Park, which provides an opportunity for Fellows to share their work and connect with the company’s broader research community. At this year’s summit in September, some of the work featured include research on the use of crowdsourcing and AI to empower citizens to counteract misinformation campaigns and teaching computers to see the world through video understanding.

Deepfake Detection Challenge and Data Set: The emergence of deepfakes and all forms of manipulated media will continue to present a challenge for society, so we are paying close attention to how research develops and are interested in working with others in the industry and academia to come up with solutions. However, we know that no single organization can solve this challenge on their own, which is why we created the first ever Deepfake Detection Challenge (DFDC) in September 2019 with our partners across industry, academia and civil society to contribute to advancing the science that detects deepfakes and manipulated media. The goal of this challenge is to produce technology that everyone can use to better detect when AI has been used to alter a video in order to mislead the viewer. As part of this initiative, we will include a dataset and leaderboard, as well as grants and awards, to spur the industry to create new ways of detecting and preventing media manipulated via AI from being used to mislead others. The governance of the challenge will be facilitated and overseen by the Partnership on AI’s new Steering Committee on AI and Media Integrity, which is made up of a broad cross-sector coalition of organizations including Facebook, WITNESS,

77 https://research.fb.com/programs/fellowship/
78 https://research.fb.com/fellows/flores-saviaga-claudia/
80 https://ai.facebook.com/blog/deepfake-detection-challenge/
Microsoft, and others in civil society and the technology, media, and academic communities. It’s important to have data that is freely available for the community to use, with clearly consenting participants, and few restrictions on usage. That’s why Facebook is commissioning a realistic data set that will use paid actors, with the required consent obtained, to contribute to the challenge. No Facebook user data will be used in this data set. We are also funding research collaborations and prizes for the challenge to help encourage more participation. In total, we are dedicating more than $10 million to fund this industry-wide effort.

**Internet Defense Prize:** The [Internet Defense Prize](https://research.fb.com/blog/2019/08/facebook-awards-100000-to-2019-internet-defense-prize-winners/) is a partnership between USENIX and Facebook that aims to reward security research that meaningfully makes the internet more secure. In August, researchers from the Max Planck Institute for Software Systems in Germany were awarded $100,000 for their work titled ERIM: Secure, Efficient In-process Isolation with Protection Keys (MPK). This research demonstrated a new approach to isolating sensitive data within software, which can help prevent a number of security issues.

**Content Policy Research on Social Media Platforms:** In May, we awarded 19 [research grants](https://research.fb.com/announcing-the-winners-of-the-content-policy-research-on-social-media-platforms-research-awards/) around the world to study our content policies, specifically around hate speech and preventing offline harm. The purpose of this research is to help us develop better content policies, assess possible interventions, and understand the mechanisms by which online rhetoric influences offline events. Grant recipients in Europe include researchers from the University of Amsterdam, Erasmus University Rotterdam, Nova University Lisbon, and King’s College London.

**Manipulated Media Research:** To stay ahead of the type of adversarial media manipulation we saw after Christchurch - when some people modified the video to avoid detection in order to repost it after it had been taken down - requires research that can be turned into technical innovation. To that end, we’ve invested in [new research partnerships](https://newsroom.fb.com/news/2019/05/protecting-live-from-abuse/) with leading academics from three universities - The University of Maryland, Cornell University and The University of California, Berkeley - to help improve image and video analysis technology. Our partners will research new techniques to 1) detect manipulated media across images, video and audio, and 2) distinguish between unwitting posters and adversaries who intentionally manipulate videos and photographs.

**Luca de Tena Foundation Grant for Research on Disinformation in Spain:** In April, in partnership with the Luca de Tena Foundation in Spain, a grant was awarded to the Universidad Rey Juan Carlos I (URJC) and the CEU San Pablo. The universities will be studying "the conditioning factors of misinformation and propose solutions against its impact depending on the degrees of vulnerability of the groups analyzed". The research is expected to be published at the end of the year.

---

82 https://research.fb.com/announcing-the-winners-of-the-content-policy-research-on-social-media-platforms-research-awards/
Appendix: Previous Facebook Code of Practice Implementation Reports:

Facebook Code of Practice on Disinformation Implementation Report | May 2019
Facebook Code of Practice on Disinformation Implementation Report | April 2019
Facebook Code of Practice on Disinformation Implementation Report | March 2019
Facebook Code of Practice on Disinformation Implementation Report | February 2019
Facebook Code of Practice on Disinformation Implementation Report | January 2019
Facebook Code of Practice on Disinformation Baseline Report