Facebook Baseline Report on Implementation of the Code of Practice on Disinformation

1. Introduction

This report provides an overview of Facebook's approach to implementing the EU Code of Practice on Disinformation, including details of our relevant policies, products, services and actions we take to address the harms caused by disinformation online. It is important to note that our approach to disinformation is in continual development, for example through the evolution of the tools we use to identify potentially false stories, clickbait and spam, and this report provides a snapshot of our approach as at January 2019. The policies, products and services detailed in this report are available globally except where we give specific details of regional coverage.

The following sections set out our current approaches to each of the categories of commitments set out in the EU Code of Practice on Disinformation.

1. Scrutiny of Ad Placements

1.1 Policies for advertising appearing on Facebook

Facebook's policies for advertising are publicly available at: https://www.facebook.com/policies/ads/. Facebook advertising policies ban the inclusion in advertising of sensational content, which we define as shocking, sensational, disrespectful or excessively violent content. We also ban the inclusion of misleading or false content: ads, landing pages, and business practices must not contain deceptive, false, or misleading content, including deceptive claims, offers, or methods.

We enforce compliance with these rules through an advertising approval process which examines the images, text, targeting, and positioning of the advertisement, in addition to the content on the advertisement's landing page. Advertisements may not be approved if the landing page content isn't fully functional, doesn't match the product/service promoted in the ad or doesn't fully comply with our Advertising Policies.

1.2 Facebook advertising network policies

Facebook's advertising network places ads on third-party sites and services, generating income for third-party publishers; Facebook policies for the advertising network also ban the inclusion of misleading, deceptive, sensational or excessively violent content. This includes deceptive claims (such as false news), offers, or business practices.

1.3 Reducing the economic incentives for false news

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

The steps we're taking to address the economic incentives for providers of false news include:

- Implementing multiple News Feed ranking changes to reduce the distribution and hence disincentivise financially-motivated tactics like the provision of clickbait, cloaking, ad farms and sharing of false or sensationalist content on the platform.
- Using signals, including feedback from people on Facebook, to predict potentially false stories for fact-checkers to review.
- Better identifying false news, drawing on feedback from our community and using third-party fact-checking organizations, so that we can limit its spread, which, in turn, makes it uneconomical. For example, when fact-checkers rate a story as false, we significantly reduce its distribution in News Feed. On average, this cuts future views by more than 80%.
- Taking action against entire Pages and websites that repeatedly share false news, reducing their overall News Feed distribution. And since we don't want to make money from misinformation or help those who create it profit, these publishers are not allowed to run ads or use our monetization features like Instant Articles.
- Applying machine learning to assist our response teams in detecting fraud and enforcing our policies against inauthentic spam accounts.
- Updating our detection of fake accounts on Facebook, which makes spamming at scale much harder.

1.4 Brand Safety

Facebook already has brand safety measures in place for ad breaks (video), Instant Articles, and Audience Network. Every piece of monetizable content is reviewed and provided a severity label for our six categories. At this time, content labeled SEVERE is ineligible to have ads placed next it to. Categories capable of attracting a SEVERE label are

- Tragedy and Conflict
- Explicit Content
- Sexual and Suggestive
- Debated Social Issues
- Objectionable Activity
- Strong Language

2. Political advertising and issue-based advertising

At Facebook we are committed to making advertising more transparent. When you visit a Facebook page or see an ad on our platform it should be clear who it is coming from. We believe that increased transparency will lead to increased accountability and responsibility. We've focused our efforts in two main areas:

- **Page Transparency**: Everywhere in the world people can now go to any page and see the ads the page is currently running. People can also see the date the page was created, any name changes it has had and any other pages that have been merged

into it. For pages with a larger following we also require the admins to authorize with us to prove they are who they say they are; we will also show the country location of those admins.

- **Political Ad Transparency**: In addition to the transparency mentioned above we also require political advertisers to take some additional steps. Anyone who wishes to run political ads must obtain authorization to do so by confirming their identity and location. They must also place a disclaimer on their ads so people know who has paid for them. Those ads go into an archive where people can see the range of impressions those ads got, the range of budget spent and the age, gender and location of who saw that ad. The ads remain in this archive for seven years. We also provide a weekly report with aggregated information about the ads in the archive.

  o **Launch Plan**: We have already launched these features in the United States, Brazil, United Kingdom and India. In the US these features cover political and issue ads. In the United Kingdom it covers political or electoral ads as well as legislation before Parliament and past referenda that are the subject of national debate, while in Brazil we only cover electoral ads. We will be launching the archive and the labelling feature, with authorisation based on an identity check, across the European Union in advance of the EU elections.

  o **News Organizations**: We have exempted news organizations from this process in the UK and plan on expanding that to other countries this year.

This transparency serves several purposes. People can see when ads are paid for by a candidate or another third-party group. It should now be more obvious when organizations are saying different things to different groups of people. In addition, journalists, watchdogs, academics, and others can use these tools to study ads on Facebook, report abuse, and hold political and issue advertisers accountable for the content they show.

3. Integrity of services

Authenticity is the cornerstone of our community and key to preserving the integrity of our services. We remove content that violates our [Community Standards](#) where we become aware of it, which are rules to ensure the safety and security of Facebook, and include explicit **requirements as to authenticity** and **prohibitions on misrepresentation**. Our authenticity and misrepresentation 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
  - Mislead people to conceal or enable the violation of other policies under the Community Standards

Our prohibition of inauthentic accounts on Facebook includes inauthentic accounts created by software (e.g., "bots").

Areas covered by these policies that have been the focus of much scrutiny and concern are fake accounts and inauthentic behavior, details of which are set out below.

3.1 Removing Fake Accounts
Fake account blocking, detection, and removal is an important aspect to preserving the integrity of Facebook's products and services. Facebook employs dedicated teams around the world to develop advanced technical systems, relying on artificial intelligence, heuristic signals, machine learning, as well as human review, to detect, block, and remove fake accounts.

Our technology helps us to take action against millions of attempts, including by bots, to create fake accounts every day, and to detect and remove millions more, often within minutes after creation. Our progress in removing fake accounts is tracked through our Community Standards Enforcement Report and select highlights from Q2 and Q3 are provided below:

- We took down more fake accounts in Q2 and Q3 2018 than in previous quarters, 800 million and 754 million, respectively. Most of these fake accounts were the result of commercially motivated spam attacks trying to create fake accounts in bulk.
  - In Q2 and Q3 2018, we found and flagged 99.6% of the accounts we subsequently took action on before users reported them. We acted on the other 0.4% because users reported them first. This number increased from 98.5% in Q1 2018.
  - Because we are able to remove most of these accounts within minutes of registration, the prevalence of fake accounts on Facebook remained steady at 3% to 4% of monthly active users as reported in our most recent (Q3 2018) earnings.
- This year we published our first Community Standards Enforcement reports, showing how much bad content we find and remove. We'll soon start releasing these reports every quarter along with conference calls, just like we do for earnings.

3.2 Prohibiting Coordinated Inauthentic Behavior

We continuously disrupt coordinated inauthentic behavior, which is when people or organizations create networks of fake accounts to mislead others about who they are, or what they're doing, to manipulate public debate for a strategic goal.

- **CIB is specifically about behavior — not content.** While we take action both against content that violates our policies and deceptive behavior, our CIB policy is designed to be behavior-based. What matters is whether the actors in question are using deceptive techniques and fake accounts. This type of content-agnostic enforcement is important, because it enables us to take action without evaluating content — or even when deceptive actors share content that would be otherwise permissible.
- Through **technical means we detect harmful activity and then flag it for manual review** by our threat intelligence and other investigative teams.
- We **take action** by having our security teams **investigate suspicious activity and take down accounts that violate our policies.**
- We **look ahead** and **work with external experts to understand the actors and risks involved.** Our partnerships include those with **governments and law enforcement, security researchers, tech industry peers, and civil society**, among other groups, and we belong to the Cybersecurity Tech Accord, a public commitment among more than 70 global companies to protect online security and defend the Internet against threats.
- Some *selected* global highlights from our takedowns for coordinated inauthentic behavior include:
    - **Belgium –** We took down 37 pages and 9 accounts around the time of the Belgian local elections, some of which were initially identified by Belgian media as potentially inauthentic and trying to manipulate political discourse, and our subsequent investigation further confirmed. Our investigation did not surface any links to foreign operators.
    - **Brazil –** We took down 68 pages and 43 accounts that were using sensationalized political content across the political spectrum to direct people to ad farms for financial gain during the Brazilian presidential election season.
    - **France -** prior to the French presidential election in 2017, we removed more than 30,000 fake accounts that were engaging in coordinated inauthentic behavior to spread spam, misinformation or other deceptive content. In removing these accounts, we identified patterns of activity, not content, that resulted in removal — for example, our systems detected repeated posting of the same content and anomalous spikes in messages sent.
    - **Iran –** We took down 104 pages, 103 accounts, 6 groups, and 92 Instagram accounts where page administrators were concealing their location and posting content focused on the Middle East, as well as the UK, U.S., and Latin America, on politically-charged topics such as race relations, opposition to the U.S. president, and immigration. Despite attempts to hide their true identity, a manual review of these accounts linked the activity to Iran.

- o **Mexico** – We took down tens of thousands of fake likes, fake pages, and fake groups to promote authentic and trustworthy civic discourse.
  - o **United States -** We took down 8 pages, 17 accounts, and 7 Instagram accounts where bad actors used VPNs and internet phone services, and paid third parties to run ads on their behalf, and some of these bad actors created an event for a protest. Inauthentic page administrators interacted with administrators of legitimate pages to co-host this event. We disabled the event, reached out to the administrators of the legitimate pages, and informed the users who were interested in the event and those who said would attend.
  - o **Myanmar** - We took down 484 pages, 157 accounts, 17 groups, and 15 Instagram where we discovered that seemingly independent news, entertainment, beauty and lifestyle pages were linked to the Myanmar military.
- As these highlights indicate, we have been proactive in detecting and removing inauthentic behavior. To stay ahead, we will continue to work collaboratively to maintain and grow this successful track record.

4. Empowering consumers

We empower people to decide for themselves what to read, trust, and share by informing them with more context in-product and promoting news literacy. For example, with the context button, we give people more details on articles and publishers. This new feature is now available to many European countries including Ireland, the UK, France, Germany, Spain and Italy. It is designed to provide people with the tools they need to make a more informed decision about which stories to read, share, and trust. Research with our community and our academic and industry partners has identified some key information that helps people evaluate the credibility of an article and determine whether to trust the article's source. Based on this research, we're making it easy for people to view context about an article, including the publisher's Wikipedia entry, related articles on the same topic, information about how many times the article has been shared on Facebook, where it is has been shared, as well as an option to follow the publisher's page. When a publisher does not have a Wikipedia entry, we will indicate that the information is unavailable, which can also be helpful context. We'll be continuing to expand coverage of EU countries as the range of available contextual information for publishers expands

When third-party fact-checkers write articles about a news story, we show them in Related Articles immediately below the story in News Feed. We also send people and Page Admins notifications if they try to share a story or have shared one in the past that's been determined to be false.

4.1 Fact-checking and false news
Facebook's fact-checking program uses a combination of technology and human review to detect and demote false news stories, which would otherwise reduce the authenticity of our service:

- In many countries Facebook is partnering with [third-party fact-checkers](#) to review and rate the accuracy of articles and posts on Facebook. These fact-checkers are independent and [certified](#) through the non-partisan International Fact-Checking Network. We use signals, including feedback from people on Facebook, to predict potentially false stories for fact-checkers to review.
- As noted in the section on *Scrutiny of Ad Placements*, we significantly reduce the distribution of stories identified as false, and Pages and domains that repeatedly share false news also see their distribution reduced and their ability to monetize and advertise removed. We use the information from fact-checkers to train our machine learning model, so that we can catch more potentially false news stories and do so faster. Finally, to give people more control, we encourage them to tell us when they see false news. [Feedback](#) from our community is one of the various signals that we use to identify potential hoaxes.
- Third party fact-checking is now available in 24 countries globally, including Denmark, France, Germany, Ireland, Italy, the Netherlands and Sweden within the EU. We will continue to learn from academics, scaling our partnerships with third-party fact-checkers and talking to other bodies like civil society organizations and journalists about how we can work together to fight misinformation.
- Any Facebook user can give us [feedback](#) that a story they're seeing in their News Feed might be false news. Feedback from our community is one of the signals that powers our machine learning model and helps us take action against stories that may be false.

4.2 Advertising transparency and consumers
The advertisements a user sees on Facebook depend on

- Information a user shares on Facebook (example: posts or comments you make) and your activity on Facebook (such as liking a Page or a post, clicking on ads you see).
- Other information about a user from their Facebook account (example: your age, your gender, your location, the devices you use to access Facebook).
- Information advertisers and our marketing partners share with Facebook that they already have, like an email address.
- User activity on websites and apps off Facebook.

The "**Why am I seeing this ad**" service, which is an option on all Facebook advertisements, provides users with an explanation of the main reasons they are seeing an ad; the service also allows users to manage their advertising experience by changing the interests relating to which they receive advertising.

4.3 Prioritising trusted sources and reducing the distribution of misleading content
In 2018, we changed News Feed to promote news from trusted sources in France, Germany, Italy, Spain and the UK. We survey diverse and representative samples of people using Facebook across the relevant markets to gauge their familiarity with, and trust in, different sources of news; and we use this data in the News Feed ranking process to promote news which is trusted by the community.

A second key pillar of our approach to prioritising trusted sources is to reduce the distribution of content which is likely to be misleading, in particular through the detection and down-ranking in News Feed of content which our users are likely to find inauthentic. As mentioned above, this reduces the economic incentives for providers of misinformation. You can learn all about how we reduce distribution of problematic content at the Facebook "Inside Feed" blog, but a few examples include:

- Clickbait: Clickbait headlines are designed to get attention and lure visitors into clicking on a link. Some headlines intentionally leave out crucial details or mislead people, forcing them to click to find out the answer. For example, "When She Looked Under Her Couch Cushions And Saw THIS…". Other headlines exaggerate the details of a story with sensational language to make the story seem like a bigger deal than it really is. For example, "WOW! Ginger tea is the secret to everlasting youth. You've GOT to see this!". We use AI tools to identify clickbait at the individual post level in addition to the domain and Page level; when we determine that a link is likely to be clickbait, we reduce its distribution in News Feed.
- Cloaking: Some providers of misleading content use a technique known as "cloaking" to circumvent Facebook's review processes and show content to people that violates Facebook's Community Standards and Advertising Policies. Here, bad actors disguise the true destination of an ad or post, or the real content of the destination page, in order to bypass Facebook's review processes. For example, they will set up web pages so that when a Facebook reviewer clicks a link to check whether it's consistent with our policies, they are taken to a different web page than when someone using the Facebook app clicks that same link. We utilize AI and human review processes to help us identify, capture, and verify cloaking - and we remove Pages that engage in cloaking.
- Ad farms: We reviewed hundreds of thousands of web pages linked to from Facebook to identify those that contain little substantive content and have a large number of disruptive, shocking or malicious ads. We use AI to assess whether new web pages shared on Facebook have similar characteristics. If we determine a post might link to these types of low-quality web pages, it will show up lower in people's News Feed and may also be determined to be ineligible to be an ad. We also downrank posts that link out to low-quality sites that predominantly copy and republish content from other sites without providing unique value.

4.4 Providing advice to voters

In addition to removing fake accounts, reducing the spread of false news and launching third party fact-checkers, we also work to provide relevant and timely information that empowers people to be informed voters in the lead up to an election. For example, in the past we've launched False News Public Service Announcements with tips on how to spot false news. We have also introduced Ballot, a voter information center that makes it easy for people to see who's running for office, follow candidate pages, and compare candidate perspectives on important issues. Candidate perspectives come directly from the candidates themselves or their staff. We provided Ballot for the recent German and Italian elections.

4.5 News Feed transparency and Inside Feed blog
We are continuing to invest in more transparency around our approach to misinformation. For example, our Inside Feed blog contains relevant, real-world examples of hoaxes that we caught and some that we didn't, as well as well as detailed explanations of our approach to fighting false news and associated issues like clickbait. See, for example https://newsroom.fb.com/news/2018/10/inside-feed-hunt-false-news-october-2018/.

4.6 Supporting media literacy and digital skills
We have worked to raise awareness of false news and boost media literacy across the EU, including a number of Member state-level projects. For example, in:

- **Germany: Media Literacy cooperation with Zeit für die Schule**
    - In 2017, we kicked off a 'school-year-long' cooperation with DIE ZEIT, aiming to increase media literacy with students 14+.
    - In Oct 2017 we launched a competition **#machdeinestory – Chefredakteure von morgen (#makeyourstory – editors in chief of tomorrow)**, which aims to motivate students to explore and tell stories, and at the same time helps them cope with the daily flood of information. We plan to give out the prizes in our very own Digitales Lernzentrum Berlin from Facebook.
    - ZEIT für die Schule and Facebook will be supporting students, helping them identifying fake news and finding reliable information. This partnership runs Oct'17-Aug'18.
- **Germany: Media Literacy cooperation with Digibits**
    - In 2016 we started our cooperation with Digibits – an NGO founded by DsiN which is supported by the Ministry of the Interior with the aim to foster online safety.
    - As part of our lasting footprint within Community Boost, in 2018 we expanded our cooperation and committed to provide funding for media literacy trainings to reach more than 100k students in Germany.
- **Italy**: **Media Literacy Campaign**
    - On the occasion of our launch of an educational tool to help people spot false news, we started a dialogue with the Ministry of Education, the Presidency of the Chamber of Deputies and other players of the industry to work together on a media literacy campaign. On May 2nd 2017, the media literacy campaign was announced during the event "#BastaBufale" (#StopHoaxes) by the former Minister of Education, Valeria Fedeli and Laura Boldrini (former President of the Chamber of Deputies].
- **United Kingdom: National Literacy Trust partnership**
    - Facebook provided funding for the National Literacy Trust's Commission on Fake News and the Teaching of Critical Literacy Skills in Schools. This report looked at how youth understand news and information on the Internet, with a primary focus on those between ages 7-11 and 11-15. The report was released in June 2018 and as a result of some of the findings, Facebook further collaborated with the National Literacy Trust to build a teachers'

resource tool that would provide access to information on digital literacy support in the classroom.
- **Poland: False News Debates**
    - In 2018 we launched a media literacy campaign in Poland called "Learning to read in the false news era". We're doing this in co-operation with [Polityka Insight](#), an independent centre for analysis and [Press](#), a key trade print media outlet in Poland.

In addition, we offer a [Digital Literacy Library](#), which has been translated into over 30 languages, including many EU languages such as Dutch, French, German, Italian, Polish and Portuguese.

5. Empowering the research community

- In April 2018, we established an independent election research commission with the goal of allowing researchers to leverage Facebook data in a privacy preserving manner to understand the impacts of our platform on Elections and Democracy. https://www.facebook.com/zuck/posts/10105865715850211

- Since April, we have worked with co-chairs Nate Persily (Stanford) and Gary King (Harvard) to establish the entity Social Science One, and build out the foundational structure of the commission in partnership with the Social Science Research Council and our foundation Funders.
- In the European Union, we established a regional advisory committee, led by [Claes Holger de Vreese](#), Professor and Chair of Political Communication in The Amsterdam School of Communication Research, University of Amsterdam. Professor de Vreese's role in the is to ensure the commission builds requests for proposals and awards research that will be valuable to the European Academic community. The European advisory commission also consists of 7 other academic representatives listed below.
    - [Marco Bastos](#), Senior Lecturer/Associate Professor at the Department of Sociology at City, University of London
    - [Frank Esser](#), Professor of International & Comparative Media Research at the University of Zurich
    - [Fabio Giglietto](#), Assistant Professor at the University of Urbino Carlo Bo
    - [Sophie Lecheler](#), Professor of Political Communication at the University of Vienna, Austria
    - [Barbara Pfetsch](#), Professor of Communication Theory and Media Effects Research at the Department of Media and Communication at the Freie Universität Berlin, Germany
    - [Cornelius Puschmann](#), Senior Researcher at the Hans Bredow Institute for Media Research in Hamburg
    - [Rebekah Tromble](#), Assistant Professor in the Institute of Political Science at Leiden University in the Netherlands

- In May 2018 we hosted a series of dinners and workshops in Oxford, Paris, and Berlin to kick off the Election Research Commission work and ensure the European academic community had a voice in the foundation structure and first datasets the commission would release.
- On September 9, 2018, the Election Research Commission hosted the Social Science One European advisory committee at Facebook's office for a European summit. The goal of the summit was to bring together the European academic and regulatory community to introduce them to the Elections Research Commission work and understand their research goals leveraging Facebook data. The whole group participated in a series of panels on the state of current social science research; the Election Research Commission Project; Facebook's approach to Elections in Europe and the current and upcoming datasets that we will be releasing around elections. We also received valuable feedback on how we could shape future data sets and RFPs to help European researchers understand the effects of our platform on democracy.
- In July 2018, we announced the first request for proposals to the research community, which includes providing researchers monetary awards as well as a dataset focused on information and misinformation shared on Facebook. The dataset consists of web page addresses (URLs) that have been shared on Facebook in the past twelve months (the dataset may grow as time passes and more URLs are shared). URLs are included if shared by many unique accounts, and shared publicly within a privacy-preserving threshold. The goal of this dataset is to allow researchers to study misinformation on Facebook and its impact on elections and democracy.
- The awards for the July 2018 RFP will be announced in January 2019, and researchers will begin receiving access to datasets in February.

- The research commission has also announced two new RFPs:

  - **Crowdtangle API**. Crowdtangle is a platform used by many media companies around the world, allowing analysts to track the popularity of news items and other public postings across multiple platforms. The Crowdtangle API will allow researchers to access both Facebook and Instagram data.

  - **Ad Archive API**. Following the launch of the ad archive in the US, we added an API to facilitate researchers' access to the data.