CODE OF PRACTICE ON DISINFORMATION

FIRST ANNUAL REPORTS – OCTOBER 2019

The Commission has received annual self-assessment reports from the online platforms and technology companies Google, Facebook, Twitter, Microsoft and Mozilla and from the trade association signatories to the Code of Practice detailing policies, processes and actions undertaken to implement their respective commitments under the Code during the Code's first year of operation.[1]

As foreseen in the December 2018 Action Plan, the Commission is now carrying out its comprehensive assessment of the effectiveness of the Code of Practice during its initial 12-month period of operation. In addition to the reporting provided by the signatories to date, the assessment will take into account the report of the Third Party Organization to be selected by the signatories, as provided for in the Code. The delivery of this report, envisioned for 15 October 2019, has been delayed.

The European Regulators Group for Audio-Visual Media Services (ERGA) is assisting the Commission in its assessment. The ERGA Sub-Group charged with this task reported in June 2019 on the first phase of its monitoring activities, which focussed on the implementation of Code commitments related to the transparency of political advertising. In December 2019, the Sub-Group will report on a second phase of its monitoring activities, which will focus at the overall effectiveness of the Code.

The Commission is being supported in its assessment by an independent consultant, which will deliver its report in January 2020. The assessment will include recommendations on next steps with respect to the Code. Should the Commission find the Code's implementation and impact have been unsatisfactory, it may propose further actions, including regulatory or co-regulatory measures.

1. OVERVIEW

The reports indicate comprehensive efforts by the signatories to implement their commitments over the last 12 months. The Code, as a self-regulatory standard, has provided an opportunity for greater transparency into the platforms' policies on disinformation as well as a framework for structured dialogue to monitor, improve and effectively implement those policies. This represents progress over the situation prevailing before the Code's entry into force, while further serious steps by individual signatories and the community as a whole are still necessary.

Reported actions taken by the platform signatories vary in terms of speed and scope across the five pillars of the Code. In general, actions to empower consumers (Pillar 4) and to empower the research community (Pillar 5) lag behind the commitments, which were subject to the Commission's targeted monitoring phase towards the European Parliament elections in May 2019. The latter concern the disruption of advertising and monetization incentives for purveyors of disinformation (Pillar 1), the transparency of political and issue-based advertising (Pillar 2), and the integrity of services against inauthentic accounts and behaviours (Pillar 3). There are also differences in the scope of actions

[1] The annual reports follow upon (i) baseline reports submitted to the Commission in January 2019 by Google, Facebook and Twitter and the European-level trade association signatories from the advertising sector setting out up-to-date information on measures taken by the end of 2018 towards implementation of the Code; and (ii) intermediate targeted monitoring reports submitted each month from January through May 2019 by Google, Facebook and Twitter focussing on the implementation of the Code's commitments with particular pertinence to the integrity of the elections.

undertaken by each platform to ensure the implementation of their commitments under the five pillars. Finally, there are differences across Member States as regards the deployment of the respective policies by the platforms.

The reports from the Code signatories also indicate some intensification of joint efforts between the platforms and other stakeholders, including fact-checkers, researchers, civil society and national authorities. These have aimed at improving the resilience of platforms' services against various forms of meddling and media manipulation and to diluting the distribution of disinformation. However, the provision of data and search tools to the research community is still episodic and arbitrary and does not respond to the full range of research needs. Moreover, cooperation with fact-checkers across the EU is still sporadic and a genuine coverage of all Member States and EU languages is still not in sight, showing the need for further efforts towards a mechanism allowing truly independent organisations to cooperate with the platforms (including via relevant and privacy-compliant access to datasets for research purposes).

Overall, the reports provide information on policies implementing the Code, including EU-specific metrics, broken down by Member States, on the enforcement of some of these policies. The consistency and level of detail in the reporting varies by platform and with respect to the five pillars.

In general, the reporting would benefit from more detail and qualitative insights in some areas (e.g. ad scrutiny, under Pillar 1) and from further big-picture context, such as trends. In addition, the metrics provided so far are mainly output indicators (e.g. number of accounts closed or ads rejected).

Finally, the reports indicate that trade association signatories have raised awareness over the past year and advocated in favour of take-up of the Code among their members. Microsoft's subscription in May 2019 extended the Code's coverage of key market participants; two additional national-level associations from the advertising sector – Association des Agences Conseils en Communication (AACC) and Stowarzyszenie Komunikacji Marketingowej/Ad Artis Art Foundation (SAR) – have also joined the Code since its entry into force. However, additional platform stakeholders have not subscribed to the Code, signalled their intention to adhere to its principles and commitments or otherwise engaged in dialogue to provide information on how they intend to resolve issues of disinformation through their services. The Commission also notes the failure so far of brands or other corporate actors from the advertising ecosystem operating in the EU to join the Code.

Set out below is a summary of the actions taken under the Code's five pillars, highlighting particular areas of progress and identifying gaps and shortcomings.

2. REPORTS OF THE SIGNATORIES

a. Scrutiny of ad placements

The platform signatories provided reporting on measures to deploy policies and processes to disrupt advertising and monetization incentives that contribute to the spread of disinformation.

Facebook reported on the application of its Advertising Policies, which prohibit a range of content, including *inter alia* sensational content, low quality or disruptive content, misleading or false content, controversial content, and the use of tactics to circumvent its ad review process or other enforcement systems. Advertising appearing on Facebook or Instagram are subject to an ad review process that

checks for compliance with these policies. In its monthly monitoring reports for March and April, Facebook reported that in each month it took action against over 600,000 ads in the EU which violated its policies on low quality or disruptive content, misleading or false content and circumvention of its systems. However, it did not provide figures on actions taken for any other months during the reporting period. In prior reporting, Facebook noted that the company reduces incentives for false news spread by online advertising, including ranking changes for its News Feed that limit the distribution of content (e.g. clickbait and false or sensational content) intended to drive traffic and garner advertising revenue. In addition, Facebook does not allow publishers which repeatedly share false information to run ads or use monetization features such as Facebook's Instant Articles. Further, in February, Facebook announced a new policy whereby the company may reject ads concerning vaccines that include misinformation.

Google reported on the enforcement of policies that prohibit advertisers across its services, and publishers participating in its AdSense network, from providing misrepresentative information about themselves, their services or their content, and from serving advertising or content which is, *inter alia*, low-quality, insufficiently original, copied and/or laden with or linked to further advertising. For the period 1 September 2018 to 31 August 2019, Google reported:

- 314,288 actions taken against EU-based Google Ads accounts for violations of its Google Ads Misrepresentation policy, and 55,876 actions for violations of its Google Ads Original Content policy; and
- 11 actions against EU-based AdSense publishers for violations of its AdSense Misrepresentative content policy, and 2,178 actions for violations of its AdSense Valuable inventory policy.

Google's reporting included break-downs of these actions by Member State. Google noted that advertisers are also subject to the Google Ads Inappropriate content policy, which prohibits content that is shocking or promotes hatred, intolerance, discrimination or violence. Google also reported on efforts in recent years to strengthen eligibility criteria for its YouTube Partner Program, which enables creators to earn revenue from ads run with their videos. Further, Google may disable advertising from running on videos that violate its Advertising-friendly content guidelines, which address content that is, *inter alia*, violent, hateful, or incendiary and demeaning.

Twitter reported on the application of its Advertising Policy , which applies to all paid advertising products – Promoted Tweets, Promoted accounts and Promoted Trends. The policy prohibits the promotion of certain broad categories of content – including hateful content, illegal products and services, and inappropriate content – as well as the promotion of unacceptable business practices, including deceptive, misleading, or harmful business propositions, misleading, false, or unsubstantiated claims, and misleading or omitted vital information on payment terms. Advertisements must also adhere to Twitter's Quality Ads Policy, which sets out editorial guidelines on *inter alia* bio and destination URLs and on content, clarity and accuracy (including a prohibition on exaggerated or sensationalist language). Advertising accounts must meet certain broad eligibility criteria and are held in review before ads may be run, and promoted content may be subject to a review process. For the period from 1 January 2019 through 31 August 2019, Twitter reported:

- 11,307 ads rejected for violation of Twitter's Unacceptable Business Practices Policies; and
- 10,639 ads rejected for violation of Twitter's Quality Ads Policy

This reporting included break-downs by Member State of the number of ads rejected. In addition, Twitter reported that in August it updated its advertising policy vis-à-vis state media, and will henceforth not accept ads from "state-controlled media".

Microsoft employs policies and enforcement processes across its services aimed at ensuring the accuracy and quality of advertising served to users. Microsoft Advertising, which serves ads displayed on Bing and provides advertising to most Microsoft services as well as to third-party services, employs policies for advertisers as well as policies for publishers that work to prevent the spread of disinformation through advertising. As regards advertisers, its Misleading Content Policies prohibit content that can be reasonably perceived as fraudulent, deceptive, or harmful to the user, including *inter alia* unsubstantiated or false claims implying affiliations or endorsements by government entities or organisations. Relevance and Quality Policies, which address advertising served through Microsoft's advertising network, prohibit advertising that leads users to landing pages or sites that employ misleading or deceptive practices or provide low-quality content or low-value user experience (e.g. misrepresentation of the origin or intent of content, high-density advertising, content that primarily supports ad monetization, phishing, malware attacks). Microsoft reported that in 2018 it suspended nearly 200,000 accounts and removed 900 million bad ads and 300,000 bad sites for violations of various of its policies. Between 1 July 2018 and 30 June 2019, it rejected more than 169 million ads as misleading. As regards publishers, Microsoft Advertising's policies include a list of prohibited content that ads may not be served against, including *inter alia*, sensitive political content (e.g. extreme, aggressive or misleading interpretation of news and events), unmoderated user-generated content, and unsavory content (e.g. disparaging individuals or organisations). Publishers must report on content management practices and Microsoft reviews publisher properties and domains for policy compliance, including compliance with restrictions on prohibited content. Separately, LinkedIn maintains Advertising Policies that prohibit fraudulent, deceptive, offensive, and incomplete ad content as well as ads for fake documents or related services. Ads on LinkedIn are subject to a review process, and users may report ads that violate the site's advertising and other policies.

In their self-assessment reports, the **trade association signatories**[2] have detailed efforts to fulfil their commitment under the Code of Practice to make their members fully aware of the Code and to encourage them to join the Code or respect its principles, as appropriate. These efforts include *inter alia* presentation and endorsement of the Code at member assemblies, discussion of the Code at meetings of relevant executive bodies, task forces and working groups, and the provision to members of informational materials concerning the Code. In addition, as called for in the Code,[3] the World Federation of Advertisers (WFA), the European Association of Communications Agencies (EACA) and IAB Europe have also provided some aggregated reporting on brand safety activities and policies employed by, respectively, brand owners, advertising agencies, and members of IAB Europe and IAB Europe's national-level associations. June 2019 saw the launch of the Global Alliance for Responsible

[2] To date, eight trade associations have subscribed to the Code of Practice. These include EDiMA, a European association representing platforms and other technology companies active in the online sector, as well as seven European-level or national-level associations representing the advertising sector: the World Federation of Advertisers (WFA), and its Belgian national member, the Union of Belgian Advertisers (UBA); the European Association of Communications Agencies (EACA), and its national members from France, Poland and the Czech Republic – respectively, Association des Agences Conseils en Communication (AACC), Stowarzyszenie Komunikacji Marketingowej /Ad Artis Art Foundation (SAR), and Asociace Komunikacnich Agentur (AKA); and the Interactive Advertising Bureau (IAB Europe).

[3] Code of Practice, Chapter III, paras. 17-19.

Media, an initiative brought forward by the WFA with the goal of *"work[ing] collaboratively to identify actions that will better protect consumers online, working towards a media environment where hate speech, bullying and disinformation is challenged …"*.

Summary

The reports indicate efforts by the platforms to disrupt advertising and monetization incentives that contribute to the dissemination of online disinformation in the EU. These include, among others, the restriction of advertising services on their platforms and the limitation of ad placements for accounts that violate the platforms' respective advertising policies and policies concerning publishers and content monetization. The platforms' enforcement of these policies work towards preventing their services from being used to spread low-quality or other abusive content across the web, including via disinformation campaigns. In addition, the platforms provided EU-specific data with respect to actions taken in the enforcement of some, but not all, of these policies.

However, the policies reported on pursue a range of objectives that are not necessarily related to the dissemination of disinformation (e.g. the promotion of misleading or unsupported commercial claims, dishonest business practices). While the platforms have reported on brand safety tools made available to advertisers on their services, there has been a lack thus far of joined-up efforts by platforms and other stakeholders – including fact-checkers and researchers and media – to identify persistent or egregious purveyors of disinformation and develop indicators for the trustworthiness of media sources, for the development and deployment of ad scrutiny and brand safety measures. The Global Alliance for Responsible Media aims at bringing together participants from across the value chain, including advertisers, ad agencies, platforms, media companies and industry associations. The Commission notes, however, that this initiative has been brought forward only recently, and that it is eager to hear about the development and deployment of concrete actions to improve transparency in the online advertising ecosystem and effectively scrutinise the ad placements. The aggregated reporting from associations in the advertising sector does not provide clarity on the extent to which brand safety practices are evolving to encompass the control of placements of advertising next to disinformation content.

b. Political advertising and issue-based advertising

All of the platform signatories provided information on their policies to ensure transparency of political ads. Only Facebook articulated a policy on issue-based advertising.

Facebook launched its political ads transparency tools globally in March 2019. Political ads on Facebook and Instagram must be clearly labelled with a "Paid for by" disclaimer. Facebook's identity confirmation and authorization system aims at preventing abuse and foreign interference. Administrators must be separately authorised in and must provide an ID document for each country which they target for political advertising. This system raised some complaints by European-wide political organisations ahead of the EU elections. Facebook's definition of political ads is wider than those of the other platforms. It covers ads made by, on behalf of or about a current or former candidate, a political party, 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 certain categories of social issues[4]; and ads regulated by law as political advertising. Between March and September 2019, Facebook served some 444,000 political ads in the EU27, totalling around 31.5 million euros of political ads spend. These figures apparently include both EU and national elections. Facebook's Ad Library is intended to contain all ads that are active and running on Facebook and Instagram, including non-political ads. Political ads remain in the library for seven years. The library shows ranges of impressions, spend and age and gender reached for each ad. The Ad Library Report provides aggregated insights on political ads, e.g. total number of ads and total spend on ads in the library by country, by advertiser or advertiser spend per day. The Ad library's API enables users to perform customized keyword searches of ads, including creatives and ad performance data.

Google published its European Union political content ads policy in January 2019, focused on the European elections, and started enforcing this policy on 21 March. It requires verification and in-ad "paid for by" disclosures for all EU election ads. Verification of advertisers running election ads includes checking their eligibility to run election ads and their identity and citizenship. The election ads policy applies to ads featuring a political party, a current elected officeholder, or a candidate for the European Parliament. Google extended this policy to national elections in the EU as of August 2019 and to national referenda as of October 2019. Between March 14 and September 16, 2019, Google received 1,541 verification applications and successfully verified 376 advertisers to run election ads in the European elections. The company labelled more than 185,000 election ad creatives. Google's Transparency Report contains 75,785 ads shown since 20 March 2019, for a total ad spend of 5,075,950 euros. Google disapproved more than 151,000 election ads where there was not proper verification of the advertiser. Google's Transparency Report on EU political advertising, launched in May 2019, offers information on who is purchasing election ads on Google and YouTube and how much money is being spent. The report includes a searchable ad library that provides information on election ads, such as the number of impressions, when ads were shown, and how ads were targeted in terms of age, gender, and location. The data from the EU Transparency Report and the ad library is also available on Google Cloud's BigQuery, which provides an API that enables unique queries.

Twitter reported on its European political campaigning ads policy, which it began enforcing in March 2019 and applied only to the European elections. For national elections, Twitter applies its global policy on political content. This global policy permits political ads in all countries but Cyprus, France, Hungary, Latvia, Lithuania and Portugal. Twitter's political ads policy for the EU elections includes a certification process, which requires that political advertisers be established in the EU in order to place political ads in any EU Member State. The objective is to ensure that only EU-based individuals can advertise political campaign content. Political ads comprise ads that fall under any of the following criteria: ads purchased by a European or national political party; ads purchased by candidates registered with their corresponding national electoral authority; and ads that advocate for or against a clearly identified candidate or party for the European elections. Twitter received 70 applications from across the EU seeking certification as political campaign accounts. As of the end of May, there were 27 certified political campaign accounts, of which 21 ran ads. These accounts accounted for some 23,253,153 impressions on Twitter and a total of 98,531 euros in revenues. Between 11 March and 26 May 2019, Twitter rejected 598 ads from non-certified accounts targeting EU Member States. Twitter also expanded its Ads Transparency Centre (ATC) to cover the EU. The ATC is a repository of all ads currently

[4] The categories covered are 1) civil and social rights; 2) economy; 3) environmental politics; 4) immigration; 5) political values and governance; and 6) security and foreign policy.

running on Twitter as well as political campaign ads run during the European elections. The ATC provides information on promoted Tweets, along with online campaign details such as billing information, ad spend, impressions data per Tweet, and demographic targeting data for ads served. Political ads will remain in the ATC indefinitely. It is unclear how consistently the ATC is applied across all Member States where political ads run.

Microsoft updated its advertising policies in April 2019 to prohibit globally ads for election-related content, political candidates, parties, ballot measures and political fundraising. This policy covers all Microsoft and third-party services that rely on Microsoft Advertising to serve advertisements. Microsoft reported that from April through August 2019 it prevented 9,703,170 political ads from serving across its ad network, 2,743,488 of which would otherwise have displayed in EU countries. Microsoft advertising policies also prohibit certain types of ads that might be considered issue-based, i.e. "advertising that exploits political agendas, sensitive political issues or uses 'hot button' political issues or names of prominent politicians […] regardless of whether the advertiser has a political agenda", as well as "advertising that exploits sensitive political or religious issues for commercial gain, or promotes extreme political or extreme religious agendas or any known associations with hate, criminal or terrorist activities."

Mozilla reported on the Mozilla Foundation's campaign for effective Ad Archive API's, which lists baseline requirements (such as the content of the advertisement and targeting criteria) that researchers need to better understand and document how disinformation spreads, including via issue-based ads.

Summary

All platform signatories deployed policies and systems to ensure transparency around political advertising, including a requirement that all political ads be clearly labelled as sponsored content and include a "paid for by" disclaimer identifying the candidate, political party or organisation paying for the ad. Although the platforms' respective definitions of political advertising are in line with the Code, there are notable differences in scope. Facebook, Google and Twitter have also reported on new authorization systems for advertisers that want to place political ads, with Facebook's being the most restrictive of the three: it requires a local authorized representative in each EU Member State where ads are targeted.

Not all the political ads served on the platforms during the EU elections were correctly labelled,[5] which diminishes the reliability of the political ads archives as well as the reporting provided on amounts spent on political advertising. The limited searchability functions of the archives as well as incomplete disclosure of targeting criteria (e.g. targeting based on users' preferences) and data about the reach of individual ads reduce the utility of the tools.

With regard to the transparency of issue-based advertising, only Facebook adopted a policy at EU level. Twitter has a specific policy and certification mechanism for issue-based ads that applies to the US only whilst in the EU issue-based ads are apparently permitted without restriction, except in France.

[5] In its report of June 2019, the European Regulators Group for Audio-Visual Media Services (ERGA) noted that Member State national regulatory authorities which participated in ERGA's intermediate monitoring of the Code found instances of "false negatives" (i.e., political ads that were not labelled as "political"). *Report of the activities carried out to assist the European Commission in the intermediate monitoring of the Code of practice on disinformation*, at pp. 13 & 16.

c. Integrity of services

All of the platform signatories reported on efforts to protect the integrity of their services against manipulative and abusive conduct.

Facebook reports preventing fake accounts by (i) blocking accounts from being created; (ii) removing accounts at sign up when they show signs of malicious behaviour; and (iii) removing accounts already on Facebook that exhibit signs of malicious behaviour. Fake accounts are identified by automated detection systems and, in the case of already active accounts, by user reports. Fake accounts blocked before creation are not included in Facebook's enforcement metrics. Facebook also reported on the enforcement of its policies against spam, understood as automated or coordinated behaviour to inflate content's distribution and reach. Facebook disabled 2.19 billion fake accounts in Q1 of 2019; updated figures on fake accounts and spam content removed in Q2 and Q3 of 2019 will not be available until November 2019. Facebook provided updates on Accounts, Pages and Groups removed for engaging in coordinated inauthentic behaviour (CIB) understood as behaviour designed to mislead about the identity of people behind an operation or the source or origin of content. Between January and October 2019, the company removed some 7,606 Accounts, Pages and Groups engaging in CIB, originating in Egypt, Honduras, Iran, Israel, Kosovo, North Macedonia, Moldova, Romania, Russia, Saudi Arabia, Spain, Thailand, the UAE, the UK and Ukraine. In recent months, Facebook has refined its policies with a view to distinguishing foreign or governmental interference as a subset of CIB and to pursuing individual inauthentic behaviour without elements of coordination.

Google's systems aim at detecting if an account creation or login is likely to be abusive, blocking such accounts and preventing other types of suspicious conduct. Google also has safeguards to detect and neutralize the impact of artificial manipulation of engagement, such as video dislikes or view counts. Google provided an update on its actions against influence operations, reporting that it had not identified any foreign-coordinated influence operations linked to the 2019 European elections. The company also reported that, between 1 September 2018 and 31 August 2019, it had removed over 10,842,500 YouTube channels for violation of its spam, misleading, and scams policy, and more than 56,500 channels for violation of its impersonation policy.

Twitter recalled its policies against malicious automation on its platform, such as commercially-motivated spam, inauthentic engagements and coordinated activity. It provided figures on actions taken to address spam, malicious automation and fake accounts: between 1 January and 31 August 2019, Twitter pro-actively challenged 126,025,294 accounts platform-wide. Twitter takes enforcement action against approximately 75% of the accounts challenged. In the same timeframe, users submitted some 4,544,096 reports of spam.

Microsoft reported on its policies to maintain the integrity of its services against threats from bots and other false account activity. In particular, it mentioned measures to prevent manipulation of Bing's search results by bots via the use of a complex ranking process based on a variety of heuristic signals. Furthermore, on LinkedIn, the use of bots or other automated methods is prohibited and a dedicated Anti-Abuse team enforces this prohibition.

Summary

The reports demonstrate that the platform signatories have policies in place to counter manipulative and inauthentic behaviour on their services. All platforms provided insights into actions undertaken to

address coordinated inauthentic behaviour, fake accounts and malicious, bot-driven activity as well as terms of service enforcement data.

However, coordinated inauthentic behaviour is still prevalent and further efforts are needed. As the figures provided are global, it is not possible to evaluate sufficiently the impact and relevance for the EU of the policies at stake. More granular information is needed to better assess malicious behaviour specifically targeting the EU and the progress achieved by the platforms to counter such behaviour. Furthermore, in order to enhance the ability of authorities to attribute conduct to specific actors, more information is needed on actions against fake accounts and the actors behind automated or human-driven malicious and inauthentic behaviour as well as companies providing artificially amplified engagement such as trading in likes, followers and shares coming from and targeting audiences in EU Member States. More detailed insights are also required about detected disinformation campaigns, including targets, levels of engagement, and the issues exploited to manipulate public opinion. Such information is needed in order to analyse whether disinformation campaigns are designed to manipulate voters in electoral contexts or, more broadly, to artificially shape public discourse around policy issues.

d. Empowering consumers

The platform signatories reported on a broad and diverse range of actions to empower consumers, including initiatives to increase media literacy, tools to better inform users, and collaborations with fact checkers.

Facebook reported on providing users with more context about the information they come across on its services, in particular by: i) reducing the distribution of disinformation in News Feed; ii) notifying users when they share content that was fact-checked and rated as "false" or "mixture", including enhanced warnings for false photos and video; iii) providing explanatory articles alongside fact-checked content; and by iv) making it easier for users to view information, via a Context Button, about websites and publishers they see on Facebook. Facebook did not report on the extent to which these features have been rolled out in all EU Member States. Additionally, Facebook has recently increased transparency for organic posts via the "Why am I seeing this post?" tool and includes a feature for users to personalize their News Feed. The company also recalled its efforts to tackle vaccine misinformation by reducing its distribution in News Feed and providing users with authoritative information on the topic. Facebook also reported on the expansion of its fact-checking collaborations, which have more than doubled in the last year. It furthermore clarified that politicians are exempted from its third-party fact-checking programme. This means that Facebook does not send organic content or ads from politicians to its third-party fact-checking partners for review. However, previously debunked content shared by politicians are demoted in News Feed.

Google recalled its collaboration with the fact-checking community and that it elevates fact-checked articles in Google News and Search. The company is also supporting initiatives to develop trustworthiness and credibility indicators for online sources (the Trust Project and the Credibility Coalition). Google reported on measures to prioritize relevant and authoritative information on its services: YouTube only surfaces content from authoritative sources on its Breaking News and Top News Shelves. These tools are currently available in 17 EU Member States. YouTube also provides context via information panels in the search results for certain historical and scientific topics that have often been subject to misinformation. However, in the EU, such contextual information is currently only

available in the UK, Germany and Spain. Furthermore, via the "Full Coverage" feature in Google News, users can access context and diverse perspectives about news stories from a variety of publishers. In September 2019, Google announced ranking updates that give more prominence to articles identified as significant original reporting, which will stay in a highly visible position for longer.

Twitter reported on the launch of a tool in May 2019 that refers users to credible public health sources when searching for keywords associated with vaccines. However, this tool is not yet rolled out in the European Union. Twitter recalled the mechanisms it makes available to users to report misleading voting-related content or potentially spammy or inauthentic accounts, and provided data on the number of user reports received. It should be noted that Twitter did not subscribe to all commitments under the Code's pillar on consumer empowerment.[6]

Microsoft reported on its "Fact Check" feature, which helps users find fact-checked content displayed within Bing search results. The company stated that its "Microsoft News" service partners with over 1,000 news sources worldwide, which are all vetted by Microsoft to ensure that the service only show licensed reputable content. It also reported on its partnership with NewsGuard, which reviews online news sites across a series of nine journalistic integrity criteria. However, the service is currently launched in only four EU Member States (Italy, Germany, France, and the UK). Furthermore, Microsoft reported on Bing's "Intelligent Search" feature, which summarizes various potentially valid answers to certain user queries and displays them in a carousel to give the user a balanced overview.

Mozilla reported on the launch of the "Firefox EU Elections Toolkit", which is a website where users can learn about tracking and opaque election advertising and how they can protect themselves. The website also includes information about the EU elections and institutions, derived from trustworthy sources. Furthermore, in September 2019, Mozilla rolled out enhanced tracking protection by default worldwide, which blocks third-party tracking cookies and crypto-miners and reduces the exposure of users to the risk of being targeted by disinformation campaigns.

All four platforms provided information on how they help consumers understand why they see particular advertisements and how they can change their preferences and control the kinds of ads they see. Facebook, Google and Twitter reported supporting a variety of media literacy campaigns and organising media literacy and digital skills trainings.

Summary

The reports demonstrate that the signatories have undertaken a variety of actions aiming at empowering consumers. In particular, all platforms provide some tools that enable consumers to understand why they are seeing particular advertisements. Likewise, all code signatories are supporting efforts to improve media literacy skills. Platforms report that they are investing in technological means to prioritize relevant, authentic and authoritative information as well as invest in features and tools that make it easier for people to find diverse perspectives about topics of public interest. However, platforms have not demonstrated much progress in developing and implementing trustworthiness indicators in collaboration with the news ecosystem. Furthermore, in general, the platforms' reporting is not detailed enough to assess the relevance and impact of the consumer

[6] Specifically, Twitter did not subscribe to commitment no. 7, which concerns *inter alia* efforts to develop trustworthiness indicators for new sources, or commitment no. 8, which concerns the prioritization of authentic and authoritative content in automatically ranked distribution channels.

empowerment tools in place; in particular, information on the uptake and actual use of these tools is lacking. Finally, some consumer empowerment tools are still not available in most EU Member States.

e. Empowering the research community

The platform signatories reported on activities to enable privacy-compliant access to data for fact-checking and research activities.

Facebook reported that in April 2018 it launched a partnership with Social Science One (SS1), a group of 83 academic researchers, to share data with the academic research community while maintaining stringent privacy protections. For this purpose, Facebook has built a custom infrastructure to provide researchers access to Facebook data in a secure manner. Personally identifiable information is removed and Facebook has tested the addition of statistical noise to ensure that individuals cannot be re-identified while simultaneously allowing for meaningful aggregate results. More than 60 researchers from 30 academic institutions across 11 countries have been chosen through a competitive peer review process organized by the Social Science Research Council (SSRC). Facebook does not play any role in this selection and has no role in directing the findings of the research. Researchers received access to data in three different ways: the CrowdTangle API, which allows researchers to access public Facebook and Instagram data related to posts from public pages, public groups and verified profiles; the Ad Library API, which provides data on ads related to politics or issues on Facebook; and an URL dataset that includes URLs that have been shared on Facebook by at least 100 unique Facebook users. In September 2019, Facebook created a Deepfake Detection Challenge with the aim of producing technology to better detect when artificial intelligence has been used to alter a video and mislead viewers.

Google reported on its partnership with the International Fact-Checking Network (IFCN), which has focused on training more fact-checkers around the world, translating the IFCN's Code of Principles into ten languages, and providing tools and training to the fact-checking community. With regard to access to Google's data for research purposes, Google recalled that its Political Ads Transparency Report, which can be downloaded as a CSV, is published as a public data set on Google Cloud BigQuery and can be searched using BigQuery's API. Google reported that the distinct daily user metrics for Google's political ads dataset are among the top 25% of all BigQuery public datasets, but did not provide concrete figures. In September 2019, Google released a large dataset on visual deepfakes that it has produced for use in developing synthetic video detection methods. The dataset has been incorporated into the FaceForensics benchmark of the Technical University of Munich and the University Federico II of Naples. Google also supported a number of research organizations like First Draft or the Oxford University's Reuters Institute with respect to issues of disinformation and trust in journalism. Google's News Lab, a team within the Google News Initiative, collaborates with journalists and entrepreneurs to drive innovation in news and hosts verification training workshops in the EU Member States.

Twitter disclosed the first comprehensive archive of state-backed information operations on Twitter in October 2018. New datasets were made available in January, June, August and September 2019, providing access to more than 30 million Tweets. Researchers in 15 EU countries accessed these datasets over 20 thousand times. Twitter's disclosure of information operations is limited to datasets associated with coordinated malicious activity that the company is able to reliably associate with state-affiliated actors. Twitter cites privacy and safety grounds for not disclosing information about individuals or accounts that are not affiliated with state actors. In addition, Twitter has been providing

an API since 2006 that, according to the company, is the largest source of real-time social media data. Twitter reported that all of its API data is public (e.g. Tweets, bios or followed accounts) and that no private user data (e.g. email addresses or IP data) is included. In addition, in January 2019, Twitter partnered with UC Berkeley to establish a new research initiative focused on improving the performance of machine learning in social systems.

Microsoft listed a number of initiatives it is working on to develop new methods and insights on how to fight disinformation. These include partnering programmes with researchers (TAP), research institutions (Princeton University, Oxford Internet Institute) and with industry, including the participation of Bing News in the Trust Project, a consortium of top news and digital companies that aims to make it easier for the public to identify quality news.

The **Mozilla** Foundation launched joint campaigns on transparency involving 71 researchers and 37 civil society organisations. The Foundation has also created a dedicated webpage with a list of tools, background reading and campaigning resources for activists and organisations working to address disinformation. In the first quarter of 2019, it completed a Fellowship Program, which supported a number of projects focusing on disinformation.

Summary

Facebook, Google and Twitter have reported on a number of policies and tools intended to provide researchers and the fact-checking community with access to platform data. This includes, in particular, access to the repositories of political ads, a resource that did not exist in the EU in 2018. The data available in these repositories is limited, in particular with regards to the targeting criteria used by political advertisers. Furthermore, while these platforms provide APIs to facilitate the running of queries by researchers and fact-checkers, there are substantial concerns about the limited functionalities of the APIs and the searchability of the repositories as a meaningful access to data for independent scrutiny, in particular in view of the ability of the platforms to alter or restrict access on a unilateral basis. Such access would enable a more rigorous monitoring and analysis of disinformation trends and impact of the measures taken by platforms to curb disinformation. Moreover, although views apparently differ among the platforms, the platforms generally cite alleged risks of data protection violations as inhibiting cooperation with the research community.

The provision of data and search tools to the research community is still episodic and arbitrary and does not respond to the full range of research needs. Moreover, a very limited community of researchers has been offered access to platforms' data sets.