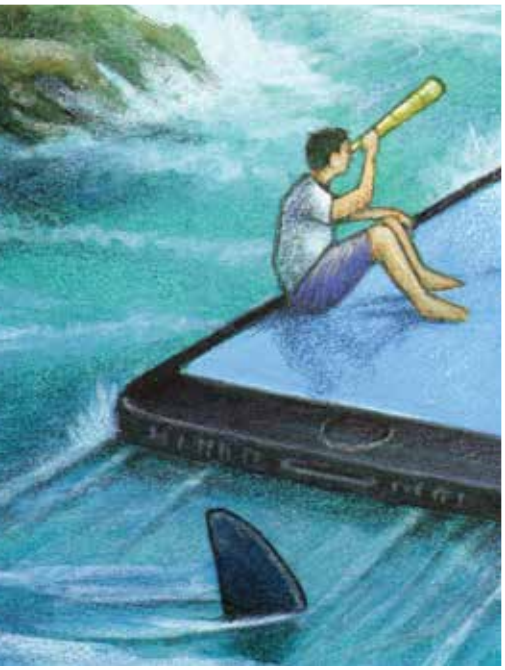
One in a Series of Working Papers from the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression



The EU Code of Practice on Disinformation:

The Difficulty of Regulating a Nebulous Problem

Peter H. Chase
The German Marshall Fund
of the United States

August 29, 2019

TRANSATLANTIC WORKING GROUP

The Transatlantic Working Group Papers Series

Co-Chairs Reports

Co-Chairs Reports from TWG's Three Sessions: Ditchley Park, Santa Monica, and Bellagio.

Freedom of Expression and Intermediary Liability

Freedom of Expression: A Comparative Summary of United States and European Law
B. Heller & J. van Hoboken, May 3, 2019.

Design Principles for Intermediary Liability Laws
J. van Hoboken & D. Keller, October 8, 2019.

Existing Legislative Initiatives

An Analysis of Germany's NetzDG Law
H. Tworek & P. Leerssen, April 15, 2019.

The Proposed EU Terrorism Content Regulation: Analysis and Recommendations with Respect to Freedom of Expression Implications
J. van Hoboken, May 3, 2019.

Combating Terrorist-Related Content Through AI and Information Sharing
B. Heller, April 26, 2019.

The European Commission's Code of Conduct for Countering Illegal Hate Speech Online: An Analysis of Freedom of Expression Implications
B. Bukovská, May 7, 2019.

The EU Code of Practice on Disinformation: The Difficulty of Regulating a Nebulous Problem
P.H. Chase, August 29, 2019.

A Cycle of Censorship: The UK White Paper on Online Harms and the Dangers of Regulating Disinformation
P. Pomerantsev, October 1, 2019.

U.S. Initiatives to Counter Harmful Speech and Disinformation on Social Media
A. Shahbaz, June 11, 2019.

ABC Framework to Address Disinformation

Actors, Behaviors, Content: A Disinformation ABC: Highlighting Three Vectors of Viral Deception to Guide Industry & Regulatory Responses
C. François, September 20, 2019.

Transparency and Accountability Solutions

Transparency Requirements for Digital Social Media Platforms: Recommendations for Policy Makers and Industry
M. MacCarthy, February 12, 2020.

Dispute Resolution and Content Moderation: Fair, Accountable, Independent, Transparent, and Effective
H. Tworek, R. Ó Fathaigh, L. Bruggeman & C. Tenove, January 14, 2020.

Algorithms and Artificial Intelligence

An Examination of the Algorithmic Accountability Act of 2019
M. MacCarthy, October 24, 2019.

Artificial Intelligence, Content Moderation, and Freedom of Expression
E. Llansó, J. van Hoboken, P. Leerssen & J. Harambam, February 26, 2020.

www.annenbergpublicpolicycenter.org/twg

A project of the Annenberg Public Policy Center of the University of Pennsylvania in partnership with
The Annenberg Foundation Trust at Sunnylands and the Institute for Information Law of the University of Amsterdam,
with support from the Embassy of the Kingdom of the Netherlands in Washington, D.C.

TRANSATLANTIC WORKING GROUP

The EU Code of Practice on Disinformation: The Difficulty of Regulating a Nebulous Problem†

Peter H. Chase, Senior Fellow, The German Marshall Fund of the United States[1]

August 29, 2019

Contents

I. Summary

The EU Code of Practice on Disinformation is a government-initiated "self-regulatory" instrument that is unlikely to achieve its goal of curtailing "disinformation." The primary hurdle the EU (and other democratic societies) faces starts with the ambiguity surrounding the concept of disinformation, which makes it difficult to define the problem and devise appropriate counter-measures. For "disinformation" points to **content** deemed to have a pernicious effect on citizens and society even though that content is not itself illegal (unlike incitement to violence or child pornography, which are caught by other laws), and regulating it directly could undermine the fundamental right to freedom of expression. To skirt around this, the Code applies only to a small group of large platforms and advertisers' associations (not publishers or other parts of the information ecosystem); contains a

† One in a series: A working paper of the Transatlantic Working Group on Content Moderation Online and Freedom of Expression. Read about the TWG: https://www.ivir.nl/twg/.

limited series of measures that may curtail advertising revenue and the impetus that gives to the dissemination of certain content; and encourages transparency, system integrity, media literacy and research access. The Code does not, however, extend to the **actors** who create the content and drive disinformation campaigns, nor does it address the inauthentic **behavior** behind the rapid and widespread dissemination of that content – two critical elements that would help narrow the problem definition to the arguably more manageable issue of "**viral deception**."

While the efforts of the social media platforms pursuant to the Code had some impact in the run-up to the May European Parliament elections, disinformation, not surprisingly, is still seen as plaguing the public debate in Europe, leading to the likelihood of the next European Commission proposing more formal regulation of the social media platforms. But this too is likely to fail, as the Commission and European politicians are unlikely to find any level of disinformation acceptable. The result could all too easily be efforts to press platforms to take down more and more "harmful" – but not illegal – content, with all the implications for freedom of expression that implies.

The report that follows provides background and political context for the creation of the Code of Conduct in Section II; describes its main provisions in Section III; notes immediate reactions to and subsequent strengthening of the Code in Section IV; summarizes actions taken by the Code signatories in Section V; reports on the Commission's evaluation of the first half-year of the Code's operation as well as some of the other critiques in Section VI; adds some additional insights about disinformation in general in Section VII; and ends with some conclusions and recommendations in Section VIII.

II. Context and Background

The EU Code of Practice on Disinformation is a specific "self-regulatory" instrument to address the problem of disinformation in the European Union. But it is only the most recent manifestation of the EU's attempts to tackle the issue, and must be seen both in the context of that evolution, as well as part of a broader program to address it.

The European Union's fight against disinformation began with Russia's sustained attacks against Estonia in 2013.[2] When the then-new Commission led by Jean-Claude Juncker entered into office at the end of 2014, the overwhelming priority of generating growth and the renewed belief in the importance of European integration led to a distinction between the awareness that defenses against foreign actors were needed and the promotion of the "Digital Single Market" for economic reasons. The first moves against disinformation accordingly were with the creation of East StratCom in the EU's External Action Service (equivalent to a State Department/Ministry of Foreign Affairs) in March 2015, specifically to counter Russian disinformation narratives.

Reflecting this initial desire to separate (foreign) "fake news" campaigns from the internet as an economic instrument, the April 2016 speech by Commission Vice President Andrus Ansip (who hails from Estonia and led the Commission on digital policy) launching the Commission's Communication on Online Platforms only mentions "fake news" in connection with false advertising, or advertising counterfeit products.

This narrative evolved as the EU entered 2017, informed by the reports of Russian interference in the U.S. presidential elections. But in his remarks to the European Parliament, Ansip takes a decidedly measured tone even at that time:

> *Fake news – or simply "lies" – are also a serious problem. We are aware of the need to protect freedom of speech and to trust people's common sense. But we also need to be aware of the possible negative effects of this phenomenon…. Self-regulation and ethical standards play a very important role here. Social media platforms and users are acting to expose fake news and unmask the source. I also see global brands and media organisations deciding to move advertising money only to sites that are known to be free from harmful content. I welcome private sector initiatives to cut commercial funding of fake news sites….*
>
> *The concept of free speech protects not only that which we agree with – but also that which is critical or disturbing. We need to address the spread of false information by improving media literacy and critical thinking, as well as by better communicating why democracy, the rule of law, protection of minorities and fundamental rights are key interests for everyone. In all these actions, we have to bear in mind that it is our responsibility to protect fundamental rights, freedom of expression in the European Union. We have to believe in the common sense of our people. Once again, fake news is bad – but Ministry of Truth is even worse.*[3]

Two months later, by mid-June 2017,[4] the tone changed as the European Parliament in its resolution on the 2016 Communication on Online Platforms stressed the need to act against the dissemination of fake news; urged platforms to supply users with tools against it; and called on the Commission to analyze current EU law on fake news and to "verify the possibility of legislative intervention to limit the dissemination and spreading of fake content."

By the end of 2017, the Commission was moving in earnest on "fake news" as something far broader (and almost divorced from) the Russian threat, announcing on November 12 that it would establish a High Level Group on Fake News, and launching a public consultation the next day during a Multi-Stakeholder Conference on Fake News. In her speech to the conference, Commissioner Mariya Gabriel stressed that the internet brings many advantages, that the EU can't revert to the days of a centralized (and usually state-owned) media, and that educating consumers to identify fake news is critical. She also laid out four key objectives – transparency, diversity of information, credibility of information, and inclusiveness. The analysis of the nearly 3,000 responses to the consultation and the delivery of the report of the High Level Group in March led directly to the Commission's April 26, 2018, Communication – Tackling Online Disinformation: A European Approach, which in turn is the basis for the Code of Practice on Disinformation, adopted in September 2018.

Perhaps because it came out barely a month following the March 17, 2018, Guardian/New York Times splash about Cambridge Analytica, the Commission's Communication took a very different tone from the High Level Group report, which was published on March 12, five days before the story hit. The High Level Group importantly succeeded in shifting the narrative away from "fake news" to disinformation (indeed, the group's name was changed to include disinformation in the title), which it also defined as:

> *Disinformation … includes all forms of false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit.*

This definition has stuck, and many of the High Level Group's recommendations are reflected in the Commission Communication. But while the High Level Group Report emphasized the importance of protecting freedom of expression, warned against hard legislation to address a "multifaceted" and rapidly evolving problem, stressed the need for evidence-based decisions and even complimented social media platforms for the many steps they had already taken to address disinformation, the Communication argued:

> *(Social media) platforms have so far failed to act proportionately, falling short of the challenge posed by disinformation and the manipulative use of platforms' infrastructures. Some have taken limited initiatives to redress the spread of online disinformation, but only in a small number of countries and leaving out many users. Furthermore, there are serious doubts whether platforms are sufficiently protecting their users against unauthorised use of their personal data by third parties, as exemplified by the recent Facebook/Cambridge Analytica revelations, currently investigated by data protection authorities, about personal data mined from millions of EU social media users and exploited in electoral contexts.*

In stark contrast to the High Level Group report, the Communication also lashed out against the social media platforms for undermining the economic viability of traditional media (whereas the High Level Group stressed that traditional media, platforms and other actors should all be part of a broad "coalition" to address disinformation), noting *inter alia* that it will use reform of the EU copyright law to "ensure a fairer distribution of revenues between rights holders and platforms, helping in particular news media outlets and journalists monetize their content."

While the remainder of this report focuses on the Code of Practice on Disinformation, which the High Level Group recommended and which the Commission then pushed, it is important to note that this is just one element of the Commission's (and European Union's) approach to disinformation, which also includes a number of other specific measures in the five areas below:

Under a "**More Transparent, Trustworthy and Accountable Online Ecosystem**:"

- the Code of Practice (below);
- strengthening fact-checking by supporting the creation of an independent network of European fact-checkers based on the International Fact-Checking Network Code of Principles and by launching a secure European online platform to support their work;
- fostering online accountability through the EU regulation on electronic identification and uptake of IPv6, which allows the allocation of a single user per internet protocol address;
- harnessing new technologies, specifically artificial intelligence, to identify, verify and tag disinformation; tools to help citizens discover disinformation; technologies to help preserve the integrity of information; and cognitive algorithms to help improve the relevance and reliability of search results;

Under a "**More Secure and Resilient Election Process**:"

- this mainly involves working with the member states to ensure the integrity of their electoral infrastructure from cyberattack;

Under "**Fostering Education and Media Literacy**:"

- working with member states on media literacy programs;
- using the Audiovisual Media Services Directive mechanisms to monitor member states' engagement in this;
- expanding the EU's own programs on digital and media literacy;
- working with the OECD to add this as a criterion in its Program for International Study Assessments (PISA);

Under "**Support for Quality Journalism as an Essential Element of a Democratic Society**:"

- facilitate member state "horizontal" support (state aids) for quality media;
- provide additional EU-level funding for initiatives promoting media freedom and pluralism, quality news media and journalism;
- promoting a toolkit for journalists on ethical issues in addressing things like disinformation from a fundamental-rights angle;

Under "**Countering Internal and External Disinformation Threats Through Strategic Communication**:"

- provide additional resources to the EU External Action Service's East StratCom Task Force, the EU Hybrid Fusion Cell and the European Centre of Excellence for Countering Hybrid Threats;
- strengthen cooperation between these EU-level organizations and member states.

III. The EU Code of Practice on Disinformation

The Code of Practice on Disinformation ("the Code") was announced by the Commission on September 26, 2018, which heralded it as the first such (government-encouraged) self-regulatory initiative in the world.[5] The Code was the product of four months of deliberation among a working group of some of the larger online platforms and advertisers, with a "Sounding Board" including other stakeholders (media, civil society, fact-checkers and academia). Facebook (including Instagram), Google (including YouTube), Mozilla and Twitter as well as four key advertising associations participated in the exercise and were the initial signatories. Microsoft joined in May 2019.

In contrast to the May 2016 EU Code of Conduct on Countering Illegal Hate Speech Online, which establishes a clear set of obligations on all participants that was explicitly negotiated with the Commission, the Commission is not so obviously associated with the Code of Practice. The Code refers to and clearly takes guidance from statements in the Commission's April Communication, but it also quickly notes that (perhaps in contrast to the issue of illegal content) the signatories all work differently, and thus have different approaches to addressing content that is not illegal. As such, not all the obligations apply equally to all signatories.

Process: As noted above, the process which led to the development of the Code involved numerous opportunities for public engagement, including in response to the Commission's Communication, the formal three-month Consultation and Eurobarometer exercise, the conference and colloquia, the engagement of the companies and organizations subject to the Code, and the Sounding Board. That said, the Code itself was never presented to the broader public for comment before being published. And, as will be noted below, the Sounding Board participants unanimously rejected the Code as inadequate.

Scope: The Code defines disinformation as "verifiably false or misleading information, which, cumulatively, is created, presented and disseminated for economic gain or to intentionally deceive the public and may cause public harm, intended as threats to democratic political and policymaking processes as well as public goods such as the protection of EU citizens' health, the environment or security." Misleading advertising, reporting errors, satire or parody, and clearly identified partisan news and commentary are explicitly ruled out of scope.[6] The Code and its commitments apply only to the territories of the countries comprising the European Economic Area (the EU plus Norway, Iceland, and Lichtenstein).

Problem Definition: The only attempt at a problem definition in the Code is noting that the signatories agree with the Commission's conclusions that "the exposure of citizens to large scale Disinformation, including misleading or outright false information, is a major challenge for Europe. Our open democratic societies depend on public debates that allow well-informed citizens to express their will through free and fair political processes."

Evidence Base: No evidence is presented in the Code to establish that disinformation causes public harm, or that the platforms that are the focus of the Code "cause" such harm. To be fair, neither is any such evidence cited in the Commission's Communication.

The only attempts in the EU's discussion to provide such evidence are in the Synopsis analyzing the 2,986 comments received during the November 2017-February 2018 public consultation as well as a Eurobarometer poll of 26,576 residents in the 28 EU member states conducted in early February 2018 (so before the March 2018 Cambridge Analytica story hit). Both are essentially opinion- rather than evidence-based, and arguably neither supports the contention. While the Eurobarometer poll establishes that well over half the respondents believe they encounter fake news nearly every day (37%) or at least once a week (31%), and nearly 85% believe fake news presents a problem, three-quarters are totally or somewhat confident they can identify it (and so presumably are not swayed by it). Further, while large majorities in most member states trust traditional media, the percentage that "totally trusts" news from online social media hovers between 1% and 3% (5% maximum), while those who "tend to trust" it is about 26%. Otherwise, European citizens take a very skeptical eye to what they see on social media. (Interestingly, there seems almost to be an inverse correlation between the two sources of news: in countries where traditional media – often state-controlled – is not trusted, social media sources are.)[7]

The Synopsis of comments received goes into more depth. It echoes the poll in terms of sources of trust in different types of media (lots for traditional sources, less for social media), perceived exposure to fake news, and a strong ability to discern it. But the more detailed questionnaire and the opportunity for open responses from the 2,784 individuals and 202 legal entities (including 69 from news media,

51 from civil society and 16 from platforms) and journalists leads the Synopsis to a more serious definition of "fake news" and the problems it causes:

> *…based on the pursued objectives of the news. The concept would mainly cover the online news, although sometimes disseminated in traditional media too, intentionally created and distributed to mislead the readers and influence their thoughts and behaviour. Fake news would seek to polarise public opinion, opinion leaders and media by creating doubts about verifiable facts, eventually jeopardising the free and democratic opinion-forming process and undermining trust in democratic processes. Gaining political or other kinds of influence, or money through online advertising (clickbait), or causing damage to an entity or a person can also be the main purpose of fake news.*

Despite the malicious nature of fake news, and the drivers perceived behind its widespread dissemination (especially on social media), the Synopsis offers no evidence of its *actual* impact (although some of the written submissions may have cited research on this); rather it states:

> *Some civil society organisations noted that both the spread and the impact of disinformation are smaller than generally assumed and that more studies are needed to properly understand the phenomenon.*

Actors: The Code applies only to the signatories – Facebook (including Instagram), Google (and YouTube), Mozilla and Twitter as well as the advertising associations that signed, specifically the European Association of Communications Agencies (EACA), the Interactive Advertising Bureau of Europe (AIB) and the World Federation of Advertisers (WFA). The latter do not enter into obligations on behalf of their members, but undertake to educate them on the Code, and to encourage them to adhere to its principles.

Objectives: "In line with the Commission Communication," the signatories agree to 11 objectives, namely, to:

- include safeguards against disinformation;
- enhance scrutiny of advertisement placements to reduce revenues to purveyors of disinformation;
- ensure transparency of political and issue-based advertising and give users means to understand why they've been targeted for it;
- implement policies against misrepresentation;
- close fake accounts and mark bots' activities to ensure they're not mistaken for human activity;
- ensure the integrity of services against accounts that spread disinformation;
- prioritize relevant, authentic and accurate information;
- ensure transparency through indicators of trustworthiness of content sources, media ownership and verified identity;
- dilute the visibility of disinformation by improving the findability of trustworthy content;

- empower users to customize newsfeeds to facilitate exposure to different views and report disinformation; and
- facilitate access to data for research.

The above is a simplified version of the actual objectives, which are more nuanced (and less onerous) than noted here. These objectives track fairly closely to the 10 principles the High Level Group recommended.

Measures: The Code underscores that, given their differences, not all signatories can work to achieve each of these objectives. As such, the signatories variously "commit" to the extent they can to adopt 15 specific measures in five categories of action:

Scrutiny of Ad Placements

- Policies and process to disrupt advertising and monetization for disinformation activities;

Political and Issue-Based Advertising

- Ensure advertised content is presented as such;
- Public disclosure of political advertising (for a candidate or on a referendum), including sponsor and amount spent;
- Public disclosure of issue-based advertising (which needs a better definition);

Integrity of Services

- Adopt clear policies regarding identity and misuse of bots;
- Adopt policies on impermissible use of bots;

Empowering Consumers

- Invest in products, technologies and programs to provide effective indicators of trustworthiness;
- Invest in technological means to prioritize trustworthy content in search, feeds or other automatically ranked distribution;
- Invest in features and tools that allow consumers to find different perspectives;
- Partner to enhance digital and media literacy;
- Help market tools to help consumers understand why they're targeted by advertising;

Empowering the Research Community

- Support independent efforts to track disinformation, including by sharing data sets and undertaking joint research;

- Don't prohibit or discourage research into disinformation and political advertising on their platforms;
- Encourage research into disinformation and political advertising;
- Convene annual meetings of stakeholders, fact-finders and researchers into these issues.

Many of the specific steps the signatories currently take in each of these areas are spelled out in an Annex of Best Practices, which provides links and details to the numerous initiatives the companies and associations had undertaken.

Remedies/Mitigation: The Code has no provisions for remedial action against unjustified takedowns of content, but neither is this a point in the Commission's Communication.

Oversight: The signatories commit to meet regularly to assess developments under the Code, to provide an annual report on activities related to the measures above, and to evaluate the effectiveness of the Code after a year, when they will discuss continuation of the Code and possible follow-up. They commit as well to engage an "objective" third party to review their self-assessments and evaluate progress toward meeting the objectives in the Code. They also commit to cooperate with the Commission, including by providing information upon request, informing the Commission of new signatories or withdrawals, responding to questions and inviting the Commission to their meetings.

IV. Immediate Reactions and Subsequent Strengthening of the Code

As noted previously, the Sounding Board[8] that was consulted as the Code was developed unanimously believed it insufficient as it contained "no common approach, no clear and meaningful commitments, no measurable objectives or KPIs (key progress indicators), hence no possibility to monitor process, and no compliance or enforcement tool."

For its part, while the Commission welcomed and indeed heralded the Code of Practice when it came out in September 2018, by December it obviously had doubts. In a December 5 Report to the European Parliament and Council on the implementation of its April Communication, the Commission notes that the Code provides an "appropriate framework" for pursuing its objectives, and disagrees with the Sounding Board by saying that the Code is consistent with Commission principles for self-regulation.

Metrics: But it also used that report (as well as an accompanying Action Plan) to effectively order the signatories to report by the end of December on actions taken, and then to report monthly through the May 2019 European elections. Furthermore, it responded to the critique of the Sounding Board by spelling out Key Progress Indicators for each of the 15 commitments (number of accounts removed for violating advertising policies, number of websites blocked for scraping content, number of political ads taken down for failing to be transparent, number of records provided to repository, number of identified fake accounts, etc.). Because the Commission would use these KPIs to measure the "success" of the Code, they are replicated in full in Appendix I.

Oversight: It also announced that it will enlist the network of member state regulators responsible for overseeing implementation of the Audio-Visual Media Services Directive (as well as the European

Audio-Visual Observatory) to assist it in monitoring compliance. Finally, the Commission signaled that "(s)hould the results prove unsatisfactory, the Commission may propose further actions, including of a regulatory nature."

V. Platform Actions Under the Code

The monthly reports the three major social media platforms that are signatories to the Code (Facebook, including Instagram; Google, including YouTube; Twitter) were compelled to issue under the Commission's Action Plan provide insight into the actions taken (in part in response to the Code but also for the companies' broader interests), and the impact those actions may have had with respect to the right to freedom of expression.

And clearly the Code had an impact. The monthly reports[9] of Facebook, Google and Twitter are structured to describe the efforts they instituted to address each of the five groups of measures that the Commission was focused on – scrutiny of ad placements; political and issue-based advertising transparency; service integrity; empowering consumers; and empowering research communities. A more detailed summary of the steps the three companies took in each of these areas is provided in Appendix 2, but some of the highlights include:

- "Google took action against 131,621 EU-based ads accounts for violating its misrepresentation policies, and against 26,824 EU-based ads accounts for violating its policies on insufficient original content; it also took action against 1,188 EU-based publisher accounts for violating its policies on valuable inventory. Facebook reported on some 1.2 million ads actioned in the EU for violating its policies on low quality or disruptive content, misleading or false content, or circumvention of its systems. Twitter reported rejecting 6,018 ads targeted at the EU for violation of its unacceptable business practices ads policy as well as 9,508 EU-targeted ads for violations of its quality ads policy;"[10]

- All three companies in March instituted new procedures to "verify" that those who want to place political ads are legitimate European-based individuals/entities, and all three by May had developed online searchable databases for these ads for all EU member states;

- Facebook also instituted similar requirements for issue-based ads related to immigration, political values, civil and social rights, security and foreign policy and environmental politics, all of which are again searchable in its Ad Library database;

- On integrity of services, Google reported taking down literally millions of YouTube channels for violating its misrepresentation and impersonation policies, while Facebook described in detail its efforts against "Coordinated Inauthentic Behavior" (including under a number of Russian-based campaigns) and noted that it took down 2.19 billion fake accounts (worldwide) during the first quarter of 2019, and Twitter reported challenging 76.6 million spam/bot/fake accounts and acting on another 2.3 million accounts reported by its users in the first five months of 2019 (again, worldwide);

- In terms of "empowering consumers," all three companies report significant efforts to support the fact-checking community, promote quality news (demote low-quality content), educate

thousands of journalists, conduct digital-media literacy campaigns for nearly a million European citizens, and help European politicians and political campaigns to protect their websites and digital communications from malicious actors;

- Finally, the companies cite efforts to help the research community, including granting access to their ads transparency websites, as well as numerous specific research projects they have funded related to disinformation, in Europe and elsewhere.

VI. Evaluating the Code

The Commission's initial evaluations of the early monthly reports submitted by Facebook, Google and Twitter were stinging. The Commission in February said it "remains deeply concerned by the failure of the platforms to identify metrics that would enable the tracking and measurement of progress in the EU as well as by lack of sufficient detail on the platforms' plans to ensure that actions in pursuit of their policies are being deployed in timely fashion and with appropriate resources across all Member States." It repeatedly "regrets" (sometimes even "deeply") that the companies did not supply sufficient information or metrics. And while in March it "takes note" of the progress the platforms made in their second monthly reports (especially on political ad transparency tools), it also stresses that "further efforts are needed by all signatories," especially in providing further metrics and details.

In its June Intermediate Targeted Monitoring evaluation of the reports the companies had submitted through the May EP elections, the Commission more positively welcomes the work of the companies, and notes:

> *The policies on transparency for online political ads implemented by the platforms as well as their actions against malicious bots, fake accounts and coordinated inauthentic behavior have likely helped limit the impact of disinformation operations from foreign and domestic actors. This is supported by a number of studies and independent sources, which suggest that the dissemination of disinformation in the run up to the European elections was not alarmingly high. For instance, according to a study by the Oxford Internet Institute, which carried out a thematic analysis of the top 20 junk news stories on Facebook and Twitter, fewer than 4% of news sources shared on Twitter ahead of the 2019 EU elections was junk news, while mainstream professional news outlets received 34% of shares. According to FactCheckEU, the European branch of IFCN, there was less disinformation than expected in the run up to the European elections and it did not dominate the conversation as it did around the past elections in Brazil, the UK, France or the United States.*

In its more formal June 14 Communication to the European Parliament and Council, the Commission reaffirms that the Code of Practice and other aspects of the Action Plan "contributed to deter attacks and expose disinformation…. raising awareness about how to counter the threat. Increased public awareness made it harder for malicious actors to manipulate the public debate." Yet the Commission acknowledges, in citing a report from Avaaz and the Institute for Strategic Dialogue, that these efforts did not stem the disinformation tide:

> *More than 600 groups and Facebook pages operating across France, Germany, Italy, the United Kingdom, Poland and Spain were reported to have spread disinformation and hate speech or have used false profiles to artificially boost the content of parties or sites they supported. These pages generated 763 million user views.*

Reports from researchers, fact-checkers and civil society also identified additional instances of large-scale attempts to manipulate voting behaviour across at least nine Member States.[11]

The Commission continues to press the platforms for additional ad transparency (particularly Google and Twitter on issue ads), more collaboration with fact-checkers and news trustworthiness indicators, and more cooperation with researchers. And again it stresses that "[s]hould the results of this assessment not be satisfactory, the Commission may propose further initiatives, including of a regulatory nature."

The Commission does not comment in its evaluations of the Code on the impact it might have had on freedom of expression. But the companies' own reports about the implementation of their "verification" process of potential political advertisers should have raised questions. By the end of May, Twitter had certified only 27 political advertising accounts; of the 676 applications Google had received, only 174 had been verified (and had run some 75,000 ads, generating €3.9 million in revenues). These numbers imply quite a lot of speech that did not benefit from amplification during the elections process, and both companies acknowledge that many of these applications were likely from legitimate sources whose applications were denied pending additional documentation. In another area, Facebook (which does not report how many political advertising accounts in Europe it did not verify) is now publishing the results of appeals of content removals that it later determined were unjustified: against the 1.1 million pieces of content removed worldwide during the first quarter of 2019 for violating Facebook's hate speech community standards, over 152,000 pieces, or 13.9 percent, were subsequently reinstated.[12] While the company's transparency on its appeals, review and reinstatement record is laudable, a more than 10 percent wrongful removal rate represents a not-insignificant impact on freedom of expression, including in the European Union.[13]

Other Commentary: Much of the other commentary about the Code published since the companies' baseline reports has criticized its voluntary nature and self-regulation in general, while stressing the importance of the long-term media-literacy efforts. A number of different sources also complain that the companies' attempts at greater transparency are still insufficient, including with respect to their political ad transparency efforts.[14] One of the more thoughtful pieces, by Paul Butcher of the European Policy Center, is more positive about the self-regulatory efforts in part as government regulation can be even more ham-fisted. He argues for greater publicity of the Code and its reports so the potential of broader public criticism (and its effect on share prices) can help hold the platforms to account, and recommends the platforms in general to be more forthcoming to such public oversight, including by civil society and researchers.

VII. Other Analyses

The EU Code of Practice on Disinformation of course was developed in the context of a much wider global debate about disinformation and its impact on broader society that goes well beyond the scope of this paper. Much of that commentary does not talk explicitly about the Code of Practice, although the analysis in it of course is pertinent to an understanding of the Code.

For instance, a recent study by the Center for the Analysis of Social Media in the British think tank Demos, Warring Songs: Information Operations in the Digital Age, provides an analysis of 39 case

studies of systemic disinformation efforts across 19 countries, including in-depth reports on those cases in three European countries.[15] The report underscores that much of the content shared in disinformation campaigns is not "fake," but selective amplification of reputable, mainstream media stories to fit an agenda. As such, it argues, the focus on fake news and disinformation is "myopic," as "information operations are vast in scale, varied in target and numerous in strategies and tactics." It goes on to define the problem more precisely as:

> *A non-kinetic, coordinated attempt to inauthentically manipulate an information environment in a systemic/strategic way, using means which are coordinated, covert and inauthentic in order to achieve political or social objectives.*

The Demos report further provides a taxonomy of the aims, strategies and tactics of such operations:

Aims, Strategies and Tactics of information operations

Aims	Strategies	Tactics
Affect sympathetic changes in behaviour and perception	Build political support Feign public support Encourage conspiratorial thinking Promote sympathetic voices	Astroturfing (fake grassroots support) False amplification of critiques of opponents False amplification of marginal voices False amplification of news Impersonation of public figures Impersonation of political allies
Reduce oppositional participation	Reduce critical voices in media Undermine trust in political representatives and institutions Undermine trust in institutions of government Undermine trust in electoral institutions Incite societal and cultural divisions Voter suppression Abuse of legal systems	Defamation Doxxing Hacking and leaking documents Interference with political processes Intimidation and harassment Dark advertising
Reduce quality of communications environment	Create confusion and anger Denigrate compromise Undermine channels of productive communication Reduce trust in digital communications Disrupt channels of communication	Exploitation of content moderation systems Playing both sides Scare stories Shocking or graphic content Communications disruption Hashtag poisoning Spam
Reduce quality of available information	Undermine trust in media institutions Undermine trust in digital media Blur the boundaries of fact and fiction Suppress critical content Promote sympathetic content Shift the balance of content in actor's favour	Algorithm exploitation and manipulations Deepfakes Dissemation of doctored images, videos and documents Dissemation of false, misleading or misattributed content Impersonation of websites Restriction of availability of information to the public Dissemation of conspiracy theories

A similar analysis of the breadth of the issue in the context of the European elections (albeit focused more on Russia as a malicious actor) highlights that "the tools and channels used to deliver disinformation to an audience will be different, and social media is not always the most important channel …. Social media platforms do not produce the malicious content; they just are used and abused to spread it. Social media may be a very powerful weapon, but the platforms are not the ones pulling the trigger."[16]

This points to an extent to a problem exacerbated by the range of actors included in the Code. For as important as the large platforms are to Europe's social discourse, they do not have monopolies –

indeed, smaller platforms can have as much (if not more) impact on sub-national/linguistic/regional polities, as can many other sources of news, including traditional media.[17]

VIII. Conclusions and Recommendations

The Code of Practice on Disinformation is a fairly messy and in some ways structurally incoherent document, but the strong political pressure behind it and the Commission's efforts to strengthen it in December 2018 with stricter reporting requirements and more oversight are clearly making a difference in the behavior of the largest platforms on advertising, system integrity, public education and research access.

While all these efforts will cut into the profits of the large platforms by reducing some advertising revenues and increasing compliance costs, they will not solve the "problem," which was poorly formulated and not well substantiated.

Disinformation by its very nature is content, in this case defined as having malicious intent – where intent is difficult to discern (although that intent can be imputed, and often is if the content isn't liked). And just as it is difficult to regulate content that is not illegal (and the Code explicitly acknowledges disinformation is not illegal), regulating only how large platforms disseminate some types of content (essentially, constraints on monetization of certain ads) will not be effective. It cannot and will not capture all malicious content (never mind "undesirable" content, which is what many politicians are concerned about); as such, it can't prevent all – or perhaps even most – of the worst instances of "viral deception."

As such, while the Code helped demonstrate the Commission was "doing something" in the run-up to the EP elections, and nudged the large platforms into better practices in a number of laudable respects, it is highly likely to be judged wanting. Indeed, the new European Commission that will enter office November 1, 2019, under the newly nominated President, Ursula von der Leyen, will propose new "hard" legislation (a "Digital Services Act") of social media platforms in part to address the problem of disinformation. This is likely to include a revision to the EU's e-Commerce Directive, which, like Section 230 under U.S. law, exempts online intermediaries from liability,[18] thus increasing compliance costs (including on small platforms that may not have the resources to make necessary technical changes, thereby increasing large platform dominance).

*****

European – and American – politicians and policy-makers are right to be concerned about the deep divisions that are appearing in their societies. They are correct as well in understanding that the rapidity with which messages spread in the online environment can exacerbate these divisions.

But they need to think carefully about their policy prescriptions to address these "harms," especially when looking at disinformation. They may not like or agree with certain content, but should bear in mind former Vice President Ansip's admonition, quoted above, that "Fake news is bad – but a Ministry of Truth is even worse."

First and most importantly, they need to distinguish between the message and its reception. If a piece of content resonates with a section of the public, whether or not that message is "factual," they need

to ask and understand why. This will not be easy, as it may point to deeper societal problems that our political systems find difficult to address. But in many ways those are the real problems, whether they are mistrust of the elites, doubts about the effectiveness of institutions (including the European Union or Congress), migration and fear of foreigners (the main themes seen in most reports about the European elections), or something else.

Second, they need to differentiate between pieces of content ("disinformation") and disruptive campaigns, that is, information operations that use the internet to generate viral deception. Here identifying the right actor is critical. Social media platforms of all sizes may be vectors for (parts of) these campaigns, but they are not the villains. Rather, those behind the campaigns are, whether they are foreign or domestic government or non-state actors using false information or selective presentation of true. Addressing the activities of those actors directly, through such efforts as the EU's StratCom Task Force or more powerful legislative acts, is arguably more important. The platforms are allies in this fight, as they need the trust of their communities and clients (advertisers) to succeed, and have the tools to disrupt at least some of the inauthentic behavior/amplification behind the campaigns. (In that sense, politicians should be as concerned about ham-handed over-removals as they are of insufficient action.)

And, as the EU's High Level Expert Group emphasized, platforms are only one part of the broader internet ecosystem that needs to be enlisted in this effort. Concerns about social media platforms taking advertising revenue from traditional media have no place in the disinformation discussion, however valid worries about the commercial health of traditional media might be.

Finally, politicians may need to admit to their publics that the "disinformation problem" cannot be resolved through self-regulatory, co-regulatory or even legislative means. This does not mean giving up. The longer-term efforts of governments, platforms and other parts of society to build media and digital literacy and to support additional research on the nature of information operations and viral deception are critical. But citizens in the end need to know that they are their own best defense, and accept that responsibility – if they are to protect their fundamental right to freedom of expression.

Appendix 1: Commission Key Performance Indicators for Code Signatories

A. Scrutiny of ad placements	
1. Deploy policies and processes to disrupt advertising and monetisation incentives for relevant behaviours	• Number of accounts removed for violation of platform advertising policies (e.g. policies against misrepresentation) • Policies put in place to demote sites or accounts that distribute disinformation or inauthentic information (e.g., click-bait) • Percentage of contracts between advertisers and ad network operators with brand safety stipulations against placement of ads on disinformation websites • Number of websites blocked for duplicating or "scraping" content produced by other websites
B. Political advertising and issue-based advertising	
2. All advertisements should be clearly distinguishable from editorial content	• Ads properly labelled as political advertising as a % of overall political ads • Actions taken to ensure all political ads are properly labelled • Number of political or issue-based ads taken down for failure to comply with platform guidelines on the transparency of political advertising
3. Enable public disclosure of political advertising	• Number of records added to public disclosure repositories • Information on amounts received from political parties, candidates, campaigns and foundations for political or issue-based advertising • Policies to verify the identity of political ads providers
4. Devising approaches to publicly disclose "issue-based advertising"	• Information on progress on this commitment
C. Integrity of services	
5. Put in place clear policies regarding identity and the misuse of automated bots on their services	• Number of identified active fake accounts • Number of identified active fake accounts disabled for violation of platform policies • Information on measures to ensure all bots are clearly labelled as such. • Number of posts, images, videos or comments acted against for violation

	of platform policies on the misuse of automated bots
6. Put in place policies on what constitutes impermissible use of automated systems	• Information on policies about the misuse of bots, including information about such bot-driven interactions • Number of bots disabled for malicious activities violating the platforms' policies

D. Empowering consumers	
7. Invest in products, technologies and programs […] to help people make informed decisions when they encounter online news that may be false	• Information on investments made in such tools or other progress towards this commitment • Information on actual use of such tools by consumers • Information on collaborations with media organisations and fact-checkers to carry out this commitment, including development of indicators of trustworthiness • Information on measures to make fact-checked content more visible and widespread.
8. Invest in technological means to prioritise relevant, authentic and authoritative information where appropriate in search, feeds, or other automatically ranked distribution channels.	• Information on progress on this commitment • Information on collaborations with media organisations and fact-checkers to carry out this commitment , including the development of indicators of trustworthiness
9. Invest in features and tools that make it easier for people to find diverse perspectives about topics of public interest	• Information on investments made in such tools or other progress towards this commitment • Information on availability of such tools and use of such tools by consumers
10. Partner with civil society, governments, educational institutions, and other stakeholders to support efforts aimed at improving critical thinking and digital media literacy	• Information about initiatives carried out or planned by signatories, including degree of coverage across Member States
11. Encourage market uptake of tools that help consumers understand why they are seeing particular advertisements	• Information on actual uptake of such tools and use by consumers

E. Empowering the research community	
12. Support good faith independent efforts to track Disinformation and understand its impact	• Information on collaborations with fact-checkers and researchers, including records shared
13. Not to prohibit or discourage good faith research into Disinformation and political advertising on their platforms	• Information on policies implementing this commitment
14. Encourage research into Disinformation and political advertising	• Information on policies implementing this commitment
15. Convene an annual event to foster discussions within academia, the fact-checking community and members of the value chain	• Report on the annual event

Appendix 2: Social Platform Actions Under the Code of Conduct

The Facebook, Google and Twitter "baseline" reports published in January 2019 are structured to allow the companies to provide more in-depth narratives about the efforts they instituted to address each of the five groups of measures noted above (ad placements, political and issue-based advertising transparency, service integrity, empowering consumers and empowering research communities), globally and within the European Union (including where certain activities are available only in some member states).

Facebook, for example, noted in its baseline report that when fact-checkers rate a story as false, it significantly reduces that story's distribution in News Feed, cutting future views by more than 80 percent. It also reported that it took down 98.5 percent more fake accounts in the second and third quarters of 2018 than in the first, 99.6 percent of which it had flagged itself (although most of these were related to commercially motivated spam). It further reported that in Belgium, it took down 37 pages and 9 accounts around the time of the local elections, some of which were initially identified by Belgian media as potentially inauthentic and trying to manipulate political discourse, as subsequent investigation further confirmed. Prior to the French presidential election in 2017, it removed more than 30,000 fake accounts that were engaging in coordinated inauthentic behavior to spread spam, misinformation or other deceptive content. Facebook also used its initial report to spell out in detail its efforts to address "Coordinated Inauthentic Behavior" (CIB), essentially content-agnostic actions to prevent the spread of content through bots and other means. **Google**, for its part, reported that in 2017 it had disapproved some 3.2 billion ads, blocked 2 million pages, terminated 320,000 publishers, and blacklisted 90,000 websites for overall content policy violations, including some 650 websites for violating its "misrepresentative content" policy. **Twitter** argued that it had made significant efforts to curb malicious automation and abuse, suspending more than 1,432,000 applications in 2018 (including 75 percent of the accounts challenged during the first half of the year). The **Mozilla** submission is shorter and specifically addresses its commitment to increase staff, roll out enhanced security features for its Firefox browser, support researchers, and launch an EP "Elections Bundle" to provide more transparency around political ad targeting. The **advertising agency** reports are primarily statements regarding efforts they have undertaken to publicize the Code among members.

In the subsequent monthly reports for January to May,[19] Facebook, Google and Twitter (the only subsequent reporters) clearly went much further. In all cases, the three platforms report on stepped-up actions against advertisers that don't meet their criteria, new processes for political (and in the case of Facebook, issue-based) ads, their focus on inauthentic behavior, and their work with politicians and the broader fact-checking and research community to find ways to detect and demote disinformation. In keeping with the Commission's emphasis in the December Report and Action Plan, the companies provide as many hard numbers as they can.

Advertising[20]

Facebook: Unlike Google and Twitter, the Facebook reports provide little statistical detail about advertisement removals on its social media platform and Instagram, beyond noting that in both March and April it "identified and actioned" over 600,000 advertisements to EU audiences that did not meet the company's standards on quality, content and/or procedures (including such things as "click-

baiting" with overly emotive images, deceptive promotion, etc.). The company argues in part that its policies and practice of reviewing ads before they are published prevents questionable ads from being shown.

More significantly, however, the company in late March launched its online Ad Library, which provides a searchable database of all ads being run on its platforms in selected countries – including, significantly, all 28 EU member states.

Google: Each of the monthly Google reports provides details (including by member state) of the number of EU-based ads and website publishers taken down for violating the company's policies on misrepresentation and content:

Issue/removals by month	January	February	March	April	May (to May 26)
Misrepresentation on Google Ads	48,642	20,627	10,234	35,428	16,690
Websites violating AdSense misrepresentation policies	0	1	0	2	0
Ads with problematic content	3,258	5,501	5,904	6,696	5,465
AdSense publishers with problematic content	205	215	370	310	88

Twitter: In the first three months, the Twitter report mainly summarizes its advertising policies; it only provides statistical data as of the April report. There, it reports that 4,590 ads that did not meet its unacceptable business-practices ad policy were prevented from being shown to European audiences during the first three months of the year, while 7,533 ads were blocked for not meeting the company's quality ads standards. The May report provides no details about April, but only about the first 20 days of the month, where the numbers are 1,428 and 1,975 respectively.

Political and Issue-Based Advertising

Facebook: Facebook in March also began its EU Political Ads process and transparency reporting, requiring verification of advertisers, labeling of ads, and transparency about their viewership. As of the end of May, there had been 343,726 political and issue ads promoted by Facebook across the EU, generating some €19 million in revenue for the company across its platforms; details for each of these ads (including about the demographics of those who viewed it) can be found on the Ad Library Report page for each of the EU member states, as well as Canada, India, Israel, Ukraine and the United States.

Google: The company brought the political ads verification process and transparency reports it had used in the United States to Europe in January 2019; the guidelines related to verifying the validity of political parties and candidates wanting to purchase ads were published in February. Subsequent reports to the Commission spell out how many applications there were to be a valid political advertiser, how many were verified/being reviewed/rejected (mainly for lack of appropriate documentation), and how many ads were approved, shown (not all approved ads by political advertisers were actually published) and rejected. The procedure for applications opened March 14, with the first labeled ads published as of March 21. As of May 28, 2019, the Google transparency report indicated that some 74,828 political ads had been shown to the European public on Google's various platforms (including YouTube) between March 21 and end May, generating €3.9 million in revenue for the company.

Google: Political Advertiser/Advertisement Data for Europe, by Month, 2019

Issue/Month	March	April	May
Advertiser applications received	120	556	676
Advertisers verified	18	123	174
Applications under review	16	13	57
Applications rejected	86	420	445
Ads approved	11,000	56,968	98,000
Ads Shown	--	10,289	63,000
Ads rejected	12,000	16,195	50,000

Twitter: Twitter's Political Advertiser Certification process and Political Ads Transparency Center began operating in Europe in March. In its May report, it notes it had received 66 applications in 11 EU member states to be certified to do political advertising in the EU; of these, 27 had been registered as of May 20. It also notes that 515 political ads (those mentioning parties or candidates in the European elections) were prevented from being shown to Europeans between March 11 and May 20 (12 of these are reported in the April report up until April 11).

Integrity of Service

Facebook: Where Twitter in many ways focused on its Political Ads Transparency initiative, Facebook's monthly submissions concentrate on its efforts to ensure the "integrity of services." This starts in part through the identification of fake accounts; in its March report, Facebook published that it had taken down 2.19 billion inauthentic accounts globally during the first quarter. It is also related to Facebook's work demoting visibility of stories that have been critiqued by fact-checkers, discussed below. More interesting is the extensive explanations Facebook provides across the reports on its work against "coordinated inauthentic behavior" on its platforms, much of which is associated with fake accounts; each month describes two to six different networks of disinformation operations that it disabled, including in Belgium, France, Kosovo, Macedonia, Moldova, Romania, Ukraine, and the United Kingdom. A substantial number of these were efforts linked to Russia, but certainly not all – in the case of Romania, for instance, the disabled networks included mainly fictitious accounts operating in support of the Social Democratic Party. Israel and Iran are identified as sources of inauthentic behavior in the May report. In this connection, Facebook announced stepping up penalties against those who abuse its CIB guidelines. Facebook also describes its work against "anti-vax" and other issues as part of its health integrity campaign.

Google: the Google reports vary in terms of their narration on "integrity of service" issues. In the baseline report and again in January, Google highlights its work under "Project Shield" in helping politicians, parties, journalists and others protect themselves against distributed denial of service (DDOS) attacks; the February reports and afterward focus more on efforts to promote quality newsfeeds on both Google News and YouTube, as well as takedowns of YouTube channels that don't meet Google policies for misrepresentation, spam, misleading content, and impersonation:

Google: YouTube Channel Removals in Europe, by Month, 2019

Reason/Month	February	March	April	May
Spam, misleading	628,000+	1,000,000+	900,000+	860,000+
Impersonation	5,000+	2,500+	500+	600+

Twitter: In its work to prevent "coordinated manipulation" in the run-up to the European Parliament elections, Twitter profited from its review of attempts to game the system during the fall 2018 U.S. elections. Like Facebook and Google, it works in part with government agencies to identify foreign interference, although voter suppression efforts were a major concern in the U.S. Twitter only began detailing numbers of bad accounts taken down in its March report:

Twitter: Accounts Challenged for Spam, Malicious Automation and Fake Accounts

Month/Source of Challenge	Proactively Challenged by Twitter (million)	Challenged by Twitter Users
January	19.5	489,148
February	17.0	406,162
March	16.6	504,729
April	13.8	597,295
May (1-20)	9.8	344,987
Total	76.7	2,342,321

The company in addition keeps an archive of potential foreign information operations, mainly pointing to Iran, Venezuela and Russia.

Electoral Support/Public Education

All three companies report on extensive efforts to promote digital media literacy in Europe, working with various civil society groups in the member states. They also all established electoral security centers, which trained candidates and political parties on ways to protect their sites from attack and abuse, as well as in reaching out to voters.

Facebook: Facebook, like the other platforms, places a lot of emphasis in its reports on partnering with fact-checkers; by its April report, it noted it had 21 fact-checking partnerships checking content in 14 European languages. In addition to its efforts to promote civic engagement (and get out the vote) and the engagement it (and all the companies had) in EU-supported Europe-wide media literacy campaigns, Facebook launched its own digital literacy campaign about "stamping out fake news" in all 28 EU member states, working with Full Fact and other fact-checking organizations; it also worked with over 20 civil society groups to conduct digital training to 75,000 citizens in seven EU countries. It reports on separate programs in countries such as Poland and Sweden, and notes that it will work with the German national newspaper Die Zeit to launch a major digital literacy program in June. It claims to have trained over 400 journalists in techniques to identify fake news stories.

Google: The company's reports highlight its efforts to protect European citizens from disinformation and to promote quality news content, mainly through Google News Lab. As of May, it had provided detailed training to nearly 6,000 European journalists in 27 of the 28 member states on news story verification techniques; helped launch FactCheck EU; provided security training to nearly 3,000

politicians and journalists; provided social media literacy training to over a million EU citizens; and promoted numerous voting, candidate and political party quick information pages.

Twitter: Among other things, Twitter has launched a new global partnership with UNESCO on media and digital literacy, which includes a series of resources to detect disinformation.

Research:

In addition to the Ad Transparency and Political Ad Transparency reports established by all three companies, Facebook and Twitter also report on other initiatives in Europe to promote and support research into the impact of social media on the public debate. Google did not spell out specific research-oriented work in Europe.

Facebook: Facebook, which in September 2018 established a European advisory committee for its Social Science One program to facilitate researcher access to its data, noted in its monthly reports that it:

- in April provided researchers from 60 universities from 30 academic institutions in 11 countries (including six EU member states) access to privacy-protected data under its Social Science One program;
- in May awarded grants for 19 research proposals on its content policies, including to four European universities;
- published in May the "audit" of the independent Data Transparency Accountability Group of its community standards and takedown activities.

Twitter:

- is actively engaging researchers to evaluate privacy and security changes to its "application program interface" (API, which is generally recognized as being relatively open for researchers);
- noted that its Potential Foreign Information Operations archive was reportedly accessed by over 13,000 researchers in Europe during the first five months of the year.

Notes

[1] Senior Fellow, German Marshall Fund of the United States. An earlier version of this paper was prepared for the May 2019 meeting of the Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression; it has been updated to reflect subsequent reporting by the signatories to the Code and revised to reflect thoughts provoked during that discussion as well as through subsequent research. The views expressed in the paper remain, however, those of the author alone.

[2] The first case of a sustained cyberattack against Estonia was in 2007, see, e.g., Damien McGuinness, How a Cyber Attack Transformed Estonia, BBC News, 27 April 2017.

[3] Statement by Commission Vice President Ansip at the European Parliament, Strasbourg, in the Plenary Debate, "Hate Speech, Populism and Fake News on Social Media – Towards an EU Response," European Commission, April 5, 2017.

[4] A little over a month following the May 7, 2017 publication in The Guardian of Carole Cadwalladr's story "The Great British Brexit Robbery: How Our Democracy was Hijacked," which reported on the role of Cambridge Analytica in the Brexit referendum.
[5] The UK government has since published an Online Harms White Paper (April 2019) which proposes an independent regulatory authority to oversee platform compliance with a "duty of care" embodied in codes of practice, including on disinformation, while the French government is also proposing a regulator to ensure platforms enforce their own terms of service/community standards.
[6] This presumably includes tweets by politicians.
[7] Hungary, Poland, Romania, Malta and Greece are all well below the EU averages in terms of trust in radio, tv, print news and online news, but above the average in terms of trust in social media and messaging aps. Eurobarometer, Fake News and Information Online, Flash Report 464, April 2018.
[8] The Opinion of the Sounding Board is available as a downloadable PDF here. Sounding Board signatories included: Grégoire Polad, Association of Commercial Television in Europe; Vincent Sneed, Association of European Radios; Oreste Pollicino, Bocconi University; Monique Goyens, Bureau Européen des Unions de Consommateurs; Ravi Vatrapu, Copenhagen Business School; Nicola Frank, European Broadcasting Union; Ricardo Gutiérrez, European Federation of Journalists; Marie de Cordier, European Magazine Media Association | European Newspaper Publishers' Association; Angela Mills Wade, European Publishers' Council; Alexios Mantzarlis, International Fact-Checking Network; Wout van Wijk, News Media Europe; Bilyana Petkova, Yale University
[9] This page on the Commission Disinformation website contains links through to all the individual monthly reports.
[10] European Commission, Joint Communication (with the EU External Action Service) to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of Regions: Report on the Implementation of the Action Plan Against Disinformation, June 14, 2019, footnote 11.
[11] Ibid. See also Alex Romero, Europe's Parliamentary Elections in the Digital Ecosystem, Disinfo Portal, July 12, 2019, updated July 15, for detailed analysis of 898 million posts across a wide range of digital media by over 95 million users in France, Germany, Italy, Poland and Spain, where one key observation is that a very small number of accounts (between 0.05 and 0.16 percent of users, mainly associated with political groups on the far left and far right, generated between 9.5 and 11 percent of activity in the five countries, mainly on socially divisive issues.
[12] Guy Rosen, An Update on How We Are Doing At Enforcing Our Community Standards, Facebook blog, May 23, 2019.
[13] For more on this problem of "over-removals," see e.g., Daphne Keller, Empirical Evidence of "Over-Removal" by Internet Companies Under Intermediary Liability Laws, Stanford University Law School Center for Internet and Society, October 12, 2015, as well "Facts and Where to Find Them: Empirical Research on Internet Platforms and Online Speech," unpublished essay, September 2018 version.
[14] See the open letter signed by 11 organizations as well as 71 researchers, Facebook and Google: This is What an Effective Ad Archive API Looks Like, The Mozilla Blog: Dispatches from the Internet frontier, March 27, 2019.
[15] Alex Krasodomski-Jones et al., Warring Songs: Information Operations in the Digital Age, Demos, Center for the Analysis of Social Media, May 2019.
[16] Jakub Kalensky, Russian Disinformation Attacks on Elections: The Case of Europe, Testimony before the Foreign Affairs Subcommittee on Europe, Eurasia, Energy and the Environment, U.S. House of Representatives, July 16, 2019, Disinfo Portal, July 17, 2019. In his testimony, Kalensky points out that the EU's East StratCom Task Force, where he previously worked, estimates that Russian disinformation activities doubled in the first half of 2019 compared to the same period the year before.
[17] See, e.g., Institute for Strategic Dialogue (ISD), Response to Online Harms White Paper Consultation, which notes, inter alia, "Harms and illegal activities are conducted through an extremely broad spectrum of technology platforms and services, as evidenced in ISD's extensive research on disinformation and extremist or terrorist use of the internet. The wide scope of platforms that would be implicated in the duty of care is, in principle, a necessity to comprehensively and sustainably address the evolving tactics of purveyors of online harm, who do not act solely on the few, largest technology platforms, but instead use an entire ecosystem of platforms to conduct harmful activity. A focus on just a few large platforms would be limited: the focus of improving content moderation approaches on a few large platforms over the past three years has led to a platform migration of many purveyors of hate speech, extremism, terrorist content and disinformation away from large platforms to smaller platforms with little or no oversight, limited or no Terms of Service (e.g. Gab), or in some cases, any appetite or intent to respond to online harms (e.g. 8chan). A limited focus on the few largest platforms would simply accelerate this phenomenon."
[18] Commission President-elect Ursula von der Leyen, A Union that Strives for More: My Agenda for Europe, European Commission, 16 July, 2019: "A new Digital Services Act will upgrade our liability and safety rules for digital platforms …" page 13.
[9] This page on the Commission Disinformation website contains links through to all the individual monthly reports.

[20] See footnote 14 supra for the online open letter published by Mozilla critiquing the transparency reports discussed here and below under "Political Advertising."