CARNEGIE
ENDOWMENT FOR
INTERNATIONAL PEACE

PARTNERSHIP FOR COUNTERING INFLUENCE OPERATIONS

*POLICY PERSPECTIVES SERIES #1*

# The EU Code of Practice on Disinformation: Briefing Note for the New EU Commission

James Pamment

MARCH 2020

# The EU Code of Practice on Disinformation: Briefing Note for the New EU Commission

James Pamment

Carnegie Endowment for International Peace
Publications Department
1779 Massachusetts Avenue NW
Washington, DC 20036
P: + 1 202 483 7600
F: + 1 202 483 1840
CarnegieEndowment.org

This publication can be downloaded at no cost at CarnegieEndowment.org.

# CONTENTS

## About the Partnership for Countering Influence Operations (PCIO)

Citizens, governments, and tech platforms around the world increasingly struggle to counter influence operations.

We believe that little progress will be made without a spirit of partnership between governments, the tech industry, media, academia, and civil society. Such collaborations are challenging but necessary in order to accomplish the three aims that PCIO believes are vital: to answer difficult policy problems related to influence operations; to find ways to understand the effect of adversarial influence operations; and to develop methods for measurement and evaluation of countermeasures.

PCIO is an international initiative, with partners and programming spanning multiple countries including in Latin America, Europe, and the Asia Pacific. PCIO and its advisory group will work actively to shape and promote an international, cross-sectoral consensus on key issues that is informed by evidence and best practice. PCIO leverages Carnegie's international networks, starting with its global centers, and is complemented by a select number of strategic partnerships. PCIO serves a convening function and as such does not speak on behalf of its members.

### About the Policy Perspectives Working Paper Series

Influence operations cannot be solved by one actor alone, yet the field is ripe with mistrust and misunderstanding between industry and government. The PCIO's Policy Perspectives working paper series offers policymakers a primer on key issues in the field while helping to build consensus among stakeholders.

## Summary


> "Progress varies a lot between signatories and the reports provide little insight on the actual impact of the self-regulatory measures taken over the past year as well as mechanisms for independent scrutiny."[1]

The EU Code of Practice on Disinformation (COP) produced mixed results. Self-regulation was a logical and necessary first step, but one year on, few of the stakeholders seem fully satisfied with the process or outcome. Strong trust has not been built between industry, governments, academia, and civil society. Most importantly, there is more to be done to better protect the public from the potential harms caused by disinformation. As with most new EU instruments, the first year of COP implementation has been difficult, and all indications are that the next year will be every bit as challenging.

This working paper offers a nonpartisan briefing on key issues for developing EU policy on disinformation. It is aimed at the incoming European Commission (EC), representatives of member states, stakeholders in the COP, and the broader community that works on identifying and countering disinformation. PCIO is an initiative of the Carnegie Endowment for International Peace and does not speak on behalf of industry or any government.

Key suggestions for the next phase include:

- **Work on cross-sector relationships.** Stop seeing each other as the problem and start building a long-term, progressive relationship to solve the real problems together.
- **Understand differences among stakeholder groups.** Member states, industry, civil society, and academia are not monolithic groups. Aim to find and build on overlapping interests where small, concrete steps can be made.
- **Focus on finding common ground.** Develop a clear vision of a future relationship among stakeholders so that all parties can plan long term to achieve this.
- **Develop a long-term collaborative focus on impact evaluation.** There are no definitive studies on the effects of either influence operations or measures to counter them, and this must be rectified as a matter of urgency.
- **Address the social media black market.** There are broader problems in how the internet is used by malign actors that can only be solved by partnerships among stakeholders.

In addition, these authors identified three key recommendations:

- **Develop a shared terminology.** The lack of common terminologies for the challenge of influence operations—among the EU and its member states and each tech platform—prevents a shared understanding of the problem, an articulation of shared goals, and instructive self-reporting on COP measures. Without agreement over a definitive EU terminological apparatus for all stakeholders to report against, opaqueness and obfuscation will continue to hamper meaningful progress.

- **Develop campaign-wide analytics for impact evaluation.** The major platforms already collaborate on intelligence sharing (including, but not limited to, attribution); in contrast to other business areas, their respective security teams have an open, trusted channel for sharing intelligence on disinformation leads and threat actor tactics, techniques, and procedures. Collaboration at the operational response level arguably indicates the feasibility of collaborating on a shared repository of analytics and campaign-wide data for policymakers and the research community.[2] This could provide an anchor point for deeper and broader multistakeholder collaboration ultimately aimed at better understanding the impact of influence operations (IO) and of countermeasures. Because nobody yet really knows what works and what does not work, the current evidence base is insufficient to support coherent policy.

- **Develop an iterative consultancy process that leads to actionable evidence.** The long-term vision should center on collaboration to develop methodologically sound research on the impact of IO and their countermeasures. PCIO supports efforts to create a sense of common purpose among diverse stakeholders, and it will launch a series of initiatives during 2020 designed to shape consensus around complex issues pertinent to the next phase of the COP.

The relationships among some COP stakeholders are fraught with tension, and there are indications that regulation derived from the Digital Services Act (DSA) and European Democracy Action Plan could complement or replace the COP.[3] At the outset of the COP, the EC had limited evidence, legal basis, political will, and terminology on which to regulate. In preparation for the next steps, whatever they may be, a more inclusive process is necessary to ensure that the regulation hits the mark. This working paper suggests some concrete steps and considerations for the road ahead.

The data used in this assessment are based on published self-reported compliance with the COP through October 2019. Therefore, the analysis is limited by the inconsistent data currently available. It is based on the first annual report, fifteen progress reports, five roadmaps, three EC evaluation reports, and the initial EC communication. Background interviews were conducted with key stakeholders and observers, and PCIO partners were given the opportunity to comment on a draft of this paper. However, the final publication is solely the responsibility of the named author.

## Background

The EC published a communication to the European Parliament in April 2018 outlining disinformation as a leading threat to democracies across Europe, as well as to the European Union (EU) itself.[4] The COP, which launched in late 2018 for a trial period of twelve months, is a central pillar of the commission's efforts to mitigate the problem.[5] With its call to act "swiftly and effectively to protect users from disinformation," the COP represents an ambitious multistakeholder approach to engaging the tech industry in a voluntary collaboration.[6] One year on, this paper examines how effective these efforts have been and outlines the key issues for the next steps on developing EU policy on disinformation.

Influence operations involve "the targeting of opinion-formation in illegitimate, though not necessarily illegal ways, by foreign actors or their proxies."[7] These campaigns use both organic and artificially generated content as tools of persuasion to influence and interfere with social relationships and political processes. Despite the difficulty of attribution in cyber operations and digital espionage, evidence suggests that the Kremlin has supported efforts to interfere with domestic politics in countries including Brazil, France, Germany, Mexico, Sweden, Ukraine, United Kingdom, United States, and Venezuela.[8]

Against this backdrop, a 2018 poll conducted by the European Commission found that 85 percent of Europeans reported thinking that "fake news" was a problem for their country and 83 percent felt that its impact on democracy in general was a problem.[9] The COP represents one subset of EU activities in this policy area. Others include the 2015 launch of the European External Action Service's East StratCom Task Force, convening of the High-Level Expert Group on Fake News and Online Disinformation, extensive efforts to protect the 2019 EU elections, launching of the Action Plan Against Disinformation, and establishment of an EU-wide Rapid Alert System.

As evaluations of the COP are finalized in early 2020, the EC has several options to take the process forward. One is to continue with a COP 2.0, with additional mechanisms and reporting requirements based on the lessons learned from the first year. A second is to continue with the self-assessment approach of the COP but to back it up with some form of regulatory intervention derived from the DSA. Such regulations would be designed to improve oversight, set minimum standards that apply to actors beyond signatories, and add some form of punishment for noncompliance; and they would also be tied to the forthcoming European Democracy Action Plan. A third option is to take a harder line on regulation, developed from the DSA. All three options raise the question of whether EU efforts should focus on the amount of verifiably false content removed or on assessing the processes and procedures used by stakeholders. Similarly, they highlight the need to address how signatory performance can be stated in useful ways that augment further private and public sector efforts to counter influence operations.

## Commitments to Progress Reporting

In October 2018, leading tech platforms voluntarily signed the COP, submitting themselves to transparent self-regulation as laid out by the EC. In January 2019, each signatory submitted a baseline report to serve as a roadmap for future efforts to combat disinformation.[10] In addition to the original requirement of a baseline report and an annual report, the commission asked in December 2018 for some signatories to submit progress reports monthly until the conclusion of the European Parliament elections in May 2019.[11] Signatory companies complied with that new request. Following the May 2019 elections, the code required signatories to publish an annual report in October 2019 on the progress of their efforts. All reports indicate progress against the following five categories of commitments outlined in the COP:

1. Scrutiny of advertisement placements
2. Political advertising and issue-based advertising
3. Integrity of services
4. Empowering consumers
5. Empowering the research community[12]

## Terminology

Many of the stakeholders involved in mitigating or countering influence operations use their own terminologies to encapsulate their view of the problem. Frequently used terms include information operations,[13] computational propaganda,[14] information manipulation,[15] information warfare,[16] information disorder,[17] hybrid warfare,[18] strategic deception,[19] and manipulative interference.[20] PCIO recommends the term *influence operations* as an umbrella term for adversary-led interference in a society; it involves techniques such as the spread of disinformation, targeted information operations, and coordinated inauthentic behavior.[21]

Terminology in this area remains challenging. Early efforts in the COP process aimed to tackle *fake news*.[22] Then in 2018, the COP switched terminology to refer to *online disinformation* and later just *disinformation*.[23] Companies also use varying terminology. For example, in their quarterly returns for the COP, Facebook referred to *coordinated inauthentic behavior*,[24] Google used terms including *influence operations* and *misrepresentation*,[25] and Twitter referred to *malicious automation*, *inauthentic activities*, and *information operations*.[26] Member states also use their own preferred terminologies, some of which provide the basis for national regulation and legislation.

Inconsistent terminology indicates a lack of consensus among key stakeholders regarding the scope of the issue and therefore its potential solutions. Clarity over objectives and terminology is required. Furthermore, the process of achieving consensus could itself inform private- and public-sector policymaking by forcing lawmakers to agree on the scope of the issue.

Meanwhile, stakeholders might continue to use their own terminologies internally to suit their view of the scope of the problem. Yet, the current method of reporting progress to the EU using ones' own terms is unsustainable. As a next step in the development of the COP, stakeholders should agree upon a definitive EU terminological apparatus for all stakeholders to report against. This should position the term *disinformation* within the broader context of operations that use it as a tactic. Terms used should be meaningful, carefully defined to support measurement and clearly linked to priority questions within this area.

## Conceptualization and Scope

The EU's High-Level Expert Group on Fake News and Online Disinformation consisted of thirty-nine experts representing online platforms, media, civil society, and academia. It focused on the roles and responsibilities of stakeholders. While such groups can suffer from a lack of focused debate, this group performed a useful role in establishing some of the key criteria that formed the COP and helped to resist the tainted term *fake news*.[27] The expert group should be reconstituted as smaller, focused, multistakeholder working groups that address specific questions at an advanced level. This could, for example, include a technical advisory group developing shared standards for technical attribution, an implementation group working with actors to verify and evaluate stakeholder activities, and a scientific development group synthesizing research developments in the field. These groups could, for example, report to the European Council's Horizontal Working Party on Enhancing Resilience and Countering Hybrid Threats.[28]

The expert group's meetings were complemented by a Eurobarometer report that polled 26,500 people across twenty-eight member states. Headline results included that 68 percent of respondents across Europe claimed to encounter fake news on a weekly basis and 71 percent were somewhat confident in their ability to identify fake news.[29] It should be noted, however, that these results are self-reported and somewhat contradictory. First, one would need to be able to identify fake news in order to know whether one was exposed to it on a weekly basis. Second, high confidence in one's own ability to identify fake news did not seem to affect the belief that fake news was a major problem for others. This is a well-established principle in social scientific research known as *third-person effect*, in which media influence is seen to affect others more than oneself—for example, in relation to video game violence.[30] Clearly, the COP requires a more rigorous and reliable research basis for

policy development. This should balance short-term needs with the long-term time frame of high-quality research. During 2020, PCIO is developing a series of briefing papers designed to orient the policy and research communities toward best practices to support researchers studying IO.

Notably, Twitter and Facebook have taken down influence operations attributed to a broader range of sources than Russia, including China, Iran, India, Pakistan, Saudi Arabia, and Venezuela. This would appear to be the tip of the iceberg since political, private sector, and civil society actors are rarely attributed publicly.[31] Furthermore, the domestic dimension of influence operations is clearly an important frontier that has not been sufficiently addressed in policymaking. The next step in the development of the COP should prioritize and define the scope of actors spreading disinformation to consider state sponsors such as China and Iran, nonstate actors such as terrorist and extremist political groups, and private sector and political actors. Clarification is also required on the germane differences between foreign and domestic influence and how these distinct but overlapping problem sets should interface with the EU's existing frameworks for hybrid and cyber policies.

## Areas of Progress

The main areas of progress are those which could be described as low-hanging fruit—that is to say, activities that do not require significant changes to an actor's business model or platform structure. Still, providing much of the requested information involved significant operational complexities, costs, and burdens for the reporting organizations. Stakeholders should also be recognized for their willingness to provide funding for strengthening the resilience of civil society to influence operations—though it is a relatively low commitment for them and shifts responsibility onto others to carry out the work. Areas of progress include:

- **Identifying and archiving political advertising.** Facebook, Google, and Twitter established Ad Libraries that catalogue political advertisers.[32] The libraries provide data on examples of the advertisers' paid content, how much money each advertiser has spent to date, the financing of such paid content, and limited insight on the impressions and engagement the advertisements received. However, representatives from civil society and research organizations have complained about the quality, depth, and breadth of information supplied by some platforms. These advances make it harder for an adversarial actor to run so-called dark ads at scale, as occurred in the 2016 U.S. presidential election. Clearly, however, more needs to be done to improve contextual data, as well as to expand it to include data on paid political speech, paid trolls, and bought engagement. In particular, industry requires clearer guidance around what is most useful to report so that they can prioritize next steps.

- **Access to takedown data.** Twitter has emerged as an industry leader in providing data from state-based IO takedowns to researchers through their Election Integrity Hub, which provides a clearinghouse for policy and data.[33] Similarly, Facebook works with third-party experts who can analyze and publish their own independent reports, including IO by nonstate actors. However, without more comprehensive access to takedown and attribution processes across the platforms (and by member states), including borderline cases, these processes can appear opaque to the wider research community. In terms of the broader interactions among industry, governments, research, and civil society, access to data remains a major hurdle to building trust and collaborating more effectively. The Twitter datasets provide an opportunity for further engagement—for example, in assessing the quality and relevance of data for research and analysis, establishing standards and mechanisms for other platforms to publish equivalent data, establishing standards and mechanisms for research and analysis, and establishing protocols for intra- and off-platform analysis. During 2020, Twitter will, in partnership with PCIO, arrange an academic conference where scholars working on the IO datasets can present their findings to Twitter data scientists.

- **Identifying and labeling misinformation.** Labeling misinformation that does not violate advertising policies has increased. For example, Facebook shows relevant articles from fact-checkers on content that's been rated *false*, *partly false*, or *false headline* by a third-party fact-checker.[34] However, it remains unclear how effective content labelling actually is, and more data should be requested to independently assess which posts are labelled, how labelling decisions are made, what various labels mean, and the impact based, for example, on AB testing.

- **Fact-checking.** Facebook and Google have made notable efforts to expand fact-checking practices. For example, Google launched the Fact Check Explorer, which links a searched term or name to independent fact-checker articles.[35] Facebook has significantly increased its work with global and regional third-party fact-checkers certified by the Poynter Institute's International Fact-Checking Network.[36] Still, there has been widespread discontent from civil society and researchers concerning the quality of monitoring tools and access to data. However, overall, fact-checking expansion is a positive development, and efforts to build civil society fact-checking capabilities should be welcomed. The EU and member states have also made significant investments in this area. More effort should be placed on developing mechanisms to ensure fact-checked data are delivered to users in a viable manner and integrated into platform structures, fostering better coordination among donors, and ensuring that data from implementation is used to better understand the impact of fact-checking on users.

- **Journalism training.** Programs such as Facebook's Journalism Project, the Google News Initiative, and the Student View are training thousands of journalists across the EU to better tell fact-supported stories, identify disinformation, and engage in high-quality digital reporting, data journalism, and visualization.[37] Journalists are key stakeholders, and efforts to build their capacity through independent civil society representatives should be welcomed. Journalists must also be educated about their role in the spread and amplification of IO, as they are often targets.

- **Media literacy.** Signatories made progress in creating and supporting media literacy programs. Facebook established a Digital Literacy Library and offered support for regional media literacy programs like Fondazione Mondo Digitale.[38] Twitter partnered with the United Nations Educational, Scientific, and Cultural Organization and provided grants to programs like Shout Out UK; Bite the Ballot; and Tea, Toast and Tweeting.[39] Google funded similar programs in Finland, France, and Portugal.[40] Media literacy programs are an essential investment in solving many of the issues related to IO over the longer term, and commitments in this area should be welcomed. Member states should also step up programming for increasing awareness for how the information environment works and what that means to end users, how they engage with information, and how this might inform their decisionmaking.

### Areas of Insufficient Progress

As the COP moves into its next phase, the focus must shift to deeper issues. The commission's first annual report expressed mixed feelings on the progress made between October 2018 and 2019.[41] While the signatories were commended for their commitment to a transparent process, the commission felt that, overall, the different parties had such divergent interests and made progress at markedly different rates and in different directions. Because the signatories face various risks, discrepancies should be expected, but more progress must be made in the following areas:

- **Defining issue-based advertising.** While Facebook identified key social issues subject to influence operations, the company's political advertisement policy currently excludes sending ads from politicians to its third-party fact-checking partners.[42] However, politicians cannot run ads that violate community standards, nor can they run ads using previously debunked content. Others running ads about social issues, elections, or politics are subject to third-party fact-checking. Twitter, on the other hand, recently introduced a blanket policy in which all political advertisements are banned.[43] Though Google has tightened its policies on political advertisements, it has expressed caution about applying censorship, and new transparency tools are anticipated in early 2020.[44] This is a telling example of different interests driving sudden policy shifts. Three different approaches are now being tested without any apparent and significant undertaking to

consistently and transparently collaborate, coordinate, or evaluate their usefulness and impact. Crucial to the next step in collaboration is consolidating these different experimental approaches into some form of harmonization—even if approaches remain platform-specific—based on any evaluation data that can be collected.

- **Detail of data.** The commission consistently urged signatories to provide more granular insights into their data. For example, Facebook reported that 2.19 billion accounts were disabled in Quarter 1 of 2019.[45] The report did not say what percentage of this 2.19 billion was commercial versus politically motivated, organic content versus bot-driven, or European-targeted or not, because many accounts were removed by automated systems and had not necessarily revealed these characteristics. Likewise, Twitter challenged 76.7 million accounts between January and May of 2019, but only 2.3 million were reported to the company by users.[46] No further insights were offered. This is in part due to the challenge of disaggregating spam from IO. While Twitter consistently identified how many ads were directed to each member state of the EU, their reports still provided limited actionable insight.

  Tech platforms argue that they are not being asked the right questions, which makes it impossible to select the most relevant data. A more collaborative, iterative process is required to ensure that platforms can fully understand and meet both the requirements and spirit of the voluntary agreement. For example, one major challenge is that the early identification and removal of fake accounts by automated systems makes it harder to develop evidence of what those accounts would end up doing. Other stakeholders must gain a better understanding of tech platforms' data collection, policy processes, and product development if they are to find a practical means of achieving the common goal of countering disinformation.

- **Evaluation of impact.** The EC has been critical of the extent to which Facebook, Twitter, and Google failed to report success metrics for their efforts. The platforms have been critical of changing requirements throughout the COP process. The companies largely reported descriptive statistics, such as the number of attendees at trainings and workshops, the amount spent on programming and grant efforts, and the number of accounts suspended or deactivated. As the commission remarked in its final statement, "reports provide little insight on the actual impact of the self-regulatory measures . . . as well as mechanisms for independent scrutiny."[47] It should be noted that none of the stakeholders (including member states and the EU) have produced comprehensive data to confirm whether their efforts are objectively and verifiably effective in either stemming disinformation, curtailing influence operations, or protecting the democratic integrity of the 2019 EU parliamentary elections.

Tech platforms have responded to data requests by collecting data on the products, policies, teams, and processes they use to mitigate IO and reporting these as metrics. There are questions about the quality of this data and how verifiable self-reported data really are. The EC and stakeholders should aim to reach clarity on what data are needed to support methodologically sound research on impact that can be used to develop policy. This is not always a question of raw data but of companies having proportionate goals, processes, products, and tools in place; this can also be evaluated. An appropriate forum to support a transparent, consultative, and iterative process is lacking.

- **Empowering research.** The commission's final statement also noted that "access to data provided so far still does not correspond to the needs of independent researchers."[48] Facebook's current model is to work closely with third-party experts to release independent analyses of takedown data for public review. Its collaboration with Social Science One encountered serious technical and governance challenges, particularly in relation to privacy concerns. Since 2006, Twitter has operated an open application program interface to access data, and it recently announced new updates to make data more accessible for researchers.[49] Twitter also maintains a one-terabyte information operations archive that contains all the IO content Twitter has removed, including account info, tweets, and media content.[50] There is frustration within the research community toward Google and Facebook in particular. However, there remains a lack of consensus over what data would be useful to research, to what ends, and under which privacy safeguards. Released data are also not used to their full potential.

  The EC should support a dialogue between research and industry to maximize opportunities for productive research aimed at better understanding the impact of IO and of countermeasures. Further, the commission should proactively examine the nature of the current regulatory environment and how it could be adjusted to enable such research, particularly in light of the challenges existing regulation has presented to Social Science One. To this end, PCIO will commission state-of-the-art reports on industry-research standards and collaboration frameworks.

## Key Issues for the Incoming Commission

### Work on the Cross-Sector Relationships

The idea that the tech platforms are somehow the problem, and that this problem can be solved by regulation, has hampered trust-building between stakeholders. The real problem is threat actors who intend to abuse the platforms, of which platforms are also victims. A regulatory response from the EU is likely to push threat actors from the mainstream platforms to other platforms that are not part of the current COP community; and while regulation can reach these actors, they are less likely to be

part of a genuinely collaborative effort. As a result, the tone and direction of future collaborations should evolve on the grounds that current signatories are still the most viable partners for developing sound EU policy. Threat actors exploiting technology are a significant problem that cannot be tackled by regulation alone. This will require significant efforts from the international community (including the EU) to advance strategies, policies, and law enforcement approaches to counter those actors. All stakeholders share a common interest in protecting end users and must learn to express this in common terms and common goals.

Regulation is also unlikely to succeed because platforms face a strategic choice when it comes to access to data. Options include, for example, establishing pragmatic structures that limit the collection of end-user data so that data requested by new regulation or legislation cannot be provided. The point is that platforms' relationships to data are a moving target. Furthermore, there remains a lack of clear thinking around how regulation might address political speech and journalistic content. Stakeholders should take a diplomatic approach to accommodating one another's needs by finding middle ground and working together on smaller issues to build trust. This should include a vision of a functional long-term relationship, grounded in the idea of measurability. The COP arguably remains the best vehicle currently available.

Internationally, the debate on regulating tech platforms is intensifying. Much of the debate seems harmful to freedom of speech, and the debate risks becoming rearticulated to suit the needs of authoritarian states. The EU should aim to become the global leader in setting reasonable, collaborative, workable, and measurable solutions to disinformation. Tech platforms and other stakeholders should make EU collaboration their priority, with a view to setting convincing global standards.

### Understand Differences Within Stakeholder Groups

It should be clear that the industry comprises as diverse a group as member states, and interests within these blocks differ. Stakeholders include web browsers, gaming platforms, cloud services, search engines, and social media platforms. Add academia and civil society to the equation, and the picture becomes highly complex and ambiguous. The EU should work to define overlapping interests—for example, in a Venn Diagram of stakeholder concerns. These points of overlap should be considered anchor points for coordinated efforts to build trust by achieving small but concrete results in narrowly focused areas. During 2020, PCIO will publish a series of working papers to support this process.

During the first year of the COP, it has become clear that different actors have different skin in the game and that they are better prepared to report on different requests based on those interests. Furthermore, the ability to quickly and efficiently produce relevant data is different for different actors. Each actor, and especially industry, needs to be able to address its own problems based on the

way its platforms and services are structured, which makes blanket regulation unlikely to succeed. Every step forward should be applauded, but ad hoc, uncoordinated measures will achieve limited results. The next phase of the COP should develop a clear strategic direction that guides actors in their long-term planning so that structures capable of meeting EC requirements can develop over time. Yet a rapidly evolving field also demands agility. Any long-term vision should be backed up by recurring check-ins and opportunities for iterative adaptations where necessary.

### Focus on Finding Common Ground

The threat of regulation inevitably leads to viewing some stakeholder relationships solely through the lens of lobbying and public policy. This fails to take advantage of the similarities between multiple areas of work all stakeholders do. Efforts should be made to broaden the scope of engagement among stakeholders, with the benefit of building trust among teams facing similar challenges and helping to shape shared solutions. For example, recent improvements in combating IO through takedowns over the past year have been driven by more effective collaboration across the industry, between civil society and industry, and between industry and government.

For example, attribution is a central element of credibly informing the public about threats from hostile actors. Currently the EU, its member states, the private sector, civil society, researchers, and tech platforms conduct technical attribution of influence operations. However, thresholds and technical standards for attribution differ among actors, making it difficult for the public (and often other parts of the community) to understand the certainty with which an attribution is made. While the tradecraft of attribution must, in some cases, remain a closely guarded secret, indicators of likelihood and the kinds of evidence used should be standardized in a way that enables the public to make a reasonably informed interpretation of a technical attribution. During 2020, PCIO will contribute to the development of community standards for various aspects of counter-IO tradecraft, including attribution.

### Develop a Long-Term Collaborative Focus on Impact Evaluation

Given the speed at which new policies and product features are announced and implemented by tech platforms, agility should be built into any future plans. However, this should not be at the expense of a clear strategy that is anchored on the principles of building and responding to an evidence base through measurement and evaluation. The next step in the development of the COP should strike a balance between pragmatism and cautious optimism. This requires clear medium- and long-term visions of the future relationship among stakeholders, of the EU's role in that relationship, and of how this relationship serves to better protect users from harm. It also requires frank dialogue among

stakeholders, as part of frequent check-ins. Though remaining agile, the strategy should focus upon the common ground of verifiable evidence—that is to say, a scientific basis for understanding the impact of IO and of countermeasures.

### Address the Social Media Black Market

Social media manipulation is not a platform-specific problem but rather intimately connected to the economics of the internet. More attention should be placed on the so-called social media black market, in which accounts, clicks, views, likes, and comments are bought and sold.[51] Accordingly, the EC should commission rigorous research into the social media black market, consider the results together with the industry, and ultimately develop remedies.

In addition, the debate surrounding the role of tech platforms as publishers is deeply flawed and risks leading to inappropriate regulation. Platforms are malleable and designed to be used in different ways by commercial and noncommercial users. Stakeholders should consider alternative methods of categorizing users, rather than overly focusing on platforms in and of themselves. For example, a user with 200 followers clearly has a different social role and responsibility than a user with 200,000 followers. YouTube has already applied a framework in this direction related to how users with broader reach monetize their accounts. Methods of enforcing different community standards, based on the reach and behavior of the user rather than the platform they use, should be further explored.[52]

Some of the failings in the COP process are symptoms of larger problems. For example, the inability of platforms to produce satisfactory data stems from a more fundamental problem of how to evaluate the impact of influence operations and countermeasures. The question of impact evaluation is the most important one and underpins many assumptions that may eventually be used to shape legislation. Therefore, it is crucial that all stakeholders work together on this foundational step. Stakeholders should place impact evaluation as the cornerstone of their future collaborative efforts, with a view to developing policy based on evaluation data. Understanding the market for social media manipulation can provide a relevant starting point that overlaps with the interests of most stakeholders.

## About the Author

**James Pamment** is a nonresident scholar in the Technology and International Affairs Program at the Carnegie Endowment for International Peace and director of the Partnership for Countering Influence Operations there. He is also an associate professor at Lund University in Sweden and the co-editor-in-chief of the Place Branding and Public Diplomacy journal. Pamment's research is about how states influence one another through strategic communication, diplomacy, intelligence, aid, and propaganda.

## Acknowledgments

The author would like to thank Phil Arceneaux and Natalie Thompson for their research support.

## Notes

1 "Code of Practice on Disinformation One Year On: Online Platforms Submit Self-Assessment Reports," European Commission, October 29, 2019, https://ec.europa.eu/commission/presscorner/detail/en/statement_19_6166.

2 It is also worth considering that trusted working relationships rely on the fact that the circle of trust is small and well-guarded, with solid legal agreements and limited risks. Building conditions of trust to scale such agreements may be challenging but is worth further consideration.

3 "New EU Rules on E-Commerce," European Commission, updated December 2, 2019, https://ec.europa.eu/digital-single-market/en/new-eu-rules-e-commerce; and Samuel Stolton and Vlagyiszlav Makszimov, "Platform Regulation 'Needed' as Part of Democracy Action Plan, Jourová Says," EURACTIV.com, updated February 5, 2020, https://www.euractiv.com/section/digital/news/platform-regulation-needed-as-part-of-democracy-action-plan-jourova-says/.

4 "Communication From the Commission to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions—Tackling Online Disinformation: A European Approach," European Commission, April 26, 2018, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0236.

5 "Code of Practice on Disinformation," European Commission, September 26, 2018, https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation.

6 "Communication From the Commission to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions."

7 James Pamment, Howard Nothhaft, Henrik Agardh-Twetman, and Alicia Fjällhed, *Countering Information Influence Activities: The State of the Art* (Stockholm, Swedish Civil Contingencies Agency, 2018), 8, https://www.msb.se/RibData/Filer/pdf/28697.pdf.

8 Robert S. Mueller, *Report On The Investigation Into Russian Interference in the 2016 Presidential Election* (Washington, DC: U.S. Department of Justice, 2019), https://www.justice.gov/storage/report.pdf; "Background to 'Assessing Russian Activities and Intentions in Recent US Elections': The Analytic Process and Cyber Incident Attribution," Office of the Director of National Intelligence, January 6, 2017, https://www.dni.gov/files/documents/ICA_2017_01.pdf; Digital, Culture, Media and Sport Committee, *Disinformation and "Fake News": Final Report* (London: House of Commons, 2019), https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf; Martin Kragh and Sebastian Åsberg, "Russia's Strategy for Influence Through Public Diplomacy and Active Measures: The Swedish Case," *Journal of Strategic Studies* 40, no. 6 (2017): 773–816, https://www.tandfonline.com/doi/full/10.1080/01402390.2016.1273830; Laura Galante and Shaun Ee, "Defining Russian Election Interference: An Analysis of Select 2014 to 2018 Cyber Enabled Incidents," Atlantic Council, September 2018, https://www.atlanticcouncil.org/wp-content/uploads/2018/09/Defining_Russian_Election_Interference_web.pdf; and Julia Gurganus, "Russia: Playing a Geopolitical Game in Latin America," Carnegie Endowment for International Peace, May 3, 2018, https://carnegieendowment.org/2018/05/03/russia-playing-geopolitical-game-in-latin-america-pub-76228.

9 Directorate-General for Communications Networks, Content & Technology, *Flash Eurobarometer 464: Fake News and Disinformation Online* (Brussels: European Commission, 2018), 4, http://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/ResultDoc/download/DocumentKy/82797.

10 "First Results of the EU Code of Practice Against Disinformation," European Commission, January 29, 2019, https://ec.europa.eu/digital-single-market/en/news/first-results-eu-code-practice-against-disinformation.

11 "A Europe That Protects: The EU Steps Up Action Against Disinformation," European Commission, December 5, 2018, https://ec.europa.eu/commission/presscorner/detail/en/IP_18_6647.

12 "EU Code of Practice on Disinformation," European Commission, September 26, 2018, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54454.

13 Yoel Roth, "Information Operations on Twitter: Principles, Process, and Disclosure," Twitter Blog, June 13, 2019, https://blog.twitter.com/en_us/topics/company/2019/information-ops-on-twitter.html.

14 Samantha Bradshaw and Philip N. Howard, *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation* (Oxford: Oxford Internet Institute, 2019), https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf.

15 Jean-Baptiste Jeangène Vilmer, Alexandre Escorcia, Marine Guillaume, and Janaina Herrera, *Information Manipulation: A Challenge for Our Democracies* (Paris: Policy Planning Staff of the Ministry for Europe and Foreign Affairs and the Institute for Strategic Research of the Ministry of the Armed Forces, 2018), https://www.diplomatie.gouv.fr/IMG/pdf/information_manipulation_rvb_cle838736.pdf.

16 Edward Lucas and Peter Pomeranzev, *Winning the Information War: Techniques and Counter-strategies to Russian Propaganda in Central and Eastern Europe* (Washington, DC: Center for European Policy Analysis, 2016), https://cepa.ecms.pl/files/?id_plik=2706.

17 Claire Wardle and Hossein Derakhshan, *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking* (Strasbourg: Council of Europe, 2017), https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html.

18 Nicole Ng and Eugene Rumer, "The West Fears Russia's Hybrid Warfare: They're Missing the Bigger Picture," Carnegie Endowment for International Peace, July 3, 2019, https://carnegieendowment.org/2019/07/03/west-fears-russia-s-hybrid-warfare.-they-re-missing-bigger-picture-pub-79412.

19 Katri Pynnöniemi and András Rácz, eds., *Fog of Falsehood: Russian Strategy of Deception and the Conflict in Ukraine* (Helsinki: Finnish Institute of International Affairs, 2016), https://www.fiia.fi/wp-content/uploads/2017/01/fiiareport45_fogoffalsehood.pdf.

20 Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, "Freedom of Expression and Elections in the Digital Age," United Nations Human Rights Special Procedures, June 2019, https://www.ohchr.org/Documents/Issues/Opinion/ElectionsReportDigitalAge.pdf.

21 Alicia Wanless and James Pamment, "How Do You Define a Problem Like Influence?" *Journal of Information Warfare* 18, no. 3 (Winter 2019): 7–8, https://carnegieendowment.org/files/2020-How_do_you_define_a_problem_like_influence.pdf.

22 "Next Steps Against Fake News: Commission Sets Up High-Level Expert Group and Launches Public Consultation," European Commission, November 13, 2017, https://ec.europa.eu/digital-single-market/en/news/next-steps-against-fake-news-commission-sets-high-level-expert-group-and-launches-public.

23 Directorate-General for Communications Networks, Content & Technology, *A Multi-Dimensional Approach to Disinformation: Report of the Independent High Level Group on Fake News and Online Disinformation* (Luxembourg: European Commission, 2018), https://op.europa.eu/en/publication-detail/-/publication/6ef4df8b-4cea-11e8-be1d-01aa75ed71a1/language-en.

24 Facebook, *Facebook Baseline Report on Implementation of the Code of Practice on Disinformation* (Brussels: European Commission, 2019), https://ec.europa.eu/information_society/newsroom/image/document/2019-5/facebook_baseline_report_on_implementation_of_the_code_of_practice_on_disinformation_CF161D11-9A54-3E27-65D58168CAC40050_56991.pdf; and "Factbox: Facebook Takedowns of 'Coordinated Inauthentic Behavior' in 2019," Reuters, August 1, 2019, https://www.reuters.com/article/us-facebook-saudi-takedowns-factbox/factbox-facebook-takedowns-of-coordinated-inauthentic-behavior-in-2019-idUSKCN1UR529.

25 Google, *EU Code of Practice on Disinformation: Google Report* (Brussels: European Commission, 2019), https://ec.europa.eu/information_society/newsroom/image/document/2019-5/google_-_ec_action_plan_reporting_CF162236-E8FB-725E-C0A3D2D6CCFE678A_56994.pdf.

26 Twitter, *Twitter Progress Report: Code of Practice on Disinformation* (Brussels: European Commission, 2019), https://ec.europa.eu/information_society/newsroom/image/document/2019-5/twitter_progress_report_on_code_of_practice_on_disinformation_CF162219-992A-B56C-06126A9E7612E13D_56993.pdf; Yoel Roth and Del Harvey, "How Twitter Is Fighting Spam and Malicious Automation," Twitter Blog, June 26, 2018, https://blog.twitter.com/en_us/topics/company/2018/how-twitter-is-fighting-spam-and-malicious-automation.html; Colin Crowell and Mahima Kaul, "Protecting the Integrity of the Election Conversation in India," Twitter Blog, February 21, 2019, https://blog.twitter.com/en_in/topics/events/2019/election-integrity.html; and Roth, "Information Operations on Twitter."

27 "Minutes of the First Meeting of the High-Level Expert Group on Fake News," European Commission, January 15, 2018, https://ec.europa.eu/information_society/newsroom/image/document/2018-6/minutes_15_january_2018_meeting_hlg_fake_news_59BB9FE9-A0B0-15BD-CB6B78C7B-820F93E_49696.pdf; and European Commission, *Synopsis Report of the Public Consultation on Fake News and Online Disinformation* (Brussels: European Commission, 2018), 7, http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51810.

28 "Horizontal Working Party on Enhancing Resilience and Countering Hybrid Threats," European Council, reviewed November 13, 2019, https://www.consilium.europa.eu/en/council-eu/preparatory-bodies/horizontal-working-party-on-enhancing-resilience-and-countering-hybrid-threats/.

29 Directorate-General for Communications Networks, Content & Technology, *Flash Eurobarometer 464*.
30 Clay Calvert, "Fake News, Censorship and the Third-Person Effect," University of Florida College of Journalism and Communications, March 13, 2017, https://www.jou.ufl.edu/insights/third-person-effect/.
31 Bradshaw and Howard, *The Global Disinformation Order*, 9.
32 Facebook, *Facebook Report on the Implementation of the Code of Practice for Disinformation: Annual Report | September 2019* (Brussels: European Commission, 2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=62681; and "Facebook Ad Library," Facebook, https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=US; Google, *EC EU Code of Practice on Disinformation: Google Annual Report* (Brussels: European Commission, 2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=62680; and "Political Advertising on Google," Google, updated February 4, 2020, https://transparencyreport.google.com/political-ads/home; and Twitter, *Twitter Progress Report: Code of Practice Against Disinformation* (Brussels: European Commission, 2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=62682; see also "Ads Transparency Center," Twitter, https://ads.twitter.com/transparency/i/political_advertisers.
33 Twitter, *Twitter Progress Report: Code of Practice against Disinformation* (Brussels: European Commission, 2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=62682; and "Elections Integrity: We're Focused on Serving the Public Conversation," Twitter, https://about.twitter.com/en_us/values/elections-integrity.html.
34 "Why Am I Seeing Related Articles Below a Post in News Feed on Facebook?" Facebook Help Center, https://www.facebook.com/help/463420517377587?helpref=faq_content.
35 Google, *EC EU Code of Practice on Disinformation*; "Fact Check Explorer," Google Fact Check Tools, https://toolbox.google.com/factcheck/explorer; and Alexios Mantzarlis, "How We Highlight Fact Checks in Search and Google News," The Keyword, December 19, 2019, https://blog.google/outreach-initiatives/google-news-initiative/how-we-highlight-fact-checks-search-and-google-news/.
36 Facebook, *Facebook Baseline Report*; and "The International Fact-Checking Network," Poynter Institute, https://www.poynter.org/ifcn/. Note that the International Fact-Checking Network does not include many established international fact-checkers.
37 Facebook, *Facebook January 2019 Update on Implementation of the Code of Practice on Disinformation* (Brussels: European Commission, 2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=57592; and "Facebook Journalism Project," Facebook Journalism Project, https://www.facebook.com/journalismproject/home; Google, *EC Action Plan on Disinformation: Google April 2019 Report* (Brussels: European Commission, 2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=59226; "Google News Initiative," Google News Initiative, https://newsinitiative.withgoogle.com/; and "Google.org Partners With Us," The Student View, https://www.thestudentview.org/updates-events/google/.
38 Facebook, *Facebook Report*; "Digital Literacy Library," Facebook, https://www.facebook.com/safety/educators; Facebook, *Facebook Reports on Implementation of the Code of Practice on Disinformation: April Report* (Brussels: European Commission, 2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=59225; and "Fondazione Mondo Digitale," Fondazione Mondo Digitale, https://www.mondodigitale.org/.

39 Twitter, *Twitter Progress Report: Code of Practice Against Disinformation* (Brussels: European Commission, 2019), http://ec.europa.eu/information_society/newsroom/image/document/2019-5/twitter_progress_report_on_code_of_practice_on_disinformation_CF162219-992A-B56C-06126A9E7612E13D_56993.pdf; Twitter, *Teaching and Learning with Twitter* (Paris: Twitter and United Nations Educational, Scientific and Cultural Organization, 2019), https://about.twitter.com/content/dam/about-twitter/values/twitter-for-good/en/teaching-learning-with-twitter-unesco.pdf; Twitter, *Twitter Progress Report: Code of Practice Against Disinformation* (Brussels: European Commission, 2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=62682; "Shout Out UK: The Voice of the Next Generation," Shout Out UK, https://www.shoutoutuk.org/; "Bite the Ballot," Bite the Ballot, https://www.bitetheballot.co.uk/; and "Tea, Toast and Tweeting," Mental Health Camden, https://mentalhealth-camden.co.uk/events/19/03/tea-toast-and-tweeting.

40 Google, *EC Action Plan on Disinformation: Google March Report* (Brussels: European Commission, 2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=58804.

41 "Code of Practice on Disinformation One Year On."

42 "Ads About Social Issues, Elections or Politics," Facebook, https://www.facebook.com/business/help/214754279118974?id=288762101909005; "10.a Ads About Social Issues, Elections or Politics," Facebook, https://www.facebook.com/policies/ads/restricted_content/political.

43 "Political Content," Twitter, https://business.twitter.com/en/help/ads-policies/prohibited-content-policies/political-content.html.

44 Scott Spencer, "An Update on our Political Ads Policy," The Keyword, November 20, 2019, https://blog.google/technology/ads/update-our-political-ads-policy.

45 "Community Standards Enforcement Report," Facebook, updated November 2019, https://transparency.facebook.com/community-standards-enforcement#fake-accounts.

46 "Platform Manipulation," Twitter, https://transparency.twitter.com/en/platform-manipulation.html.

47 "Code of Practice on Disinformation One Year On."

48 Ibid.

49 "Twitter Data for Academic Research," Twitter, https://developer.twitter.com/en/use-cases/academic-researchers; and "Helpful Tools: Resources for Researchers," Twitter, https://developer.twitter.com/en/use-cases/academic-researchers/helpful-tools.

50 "Information Operations," Twitter, https://transparency.twitter.com/en/information-operations.html.

51 Sebastian Bay and Rolf Fredheim, *Falling Behind: How Social Media Companies are Failing to Combat Inauthentic Behaviour Online* (Riga: NATO Strategic Communications Centre of Excellence, 2019), 6, https://www.stratcomcoe.org/how-social-media-companies-are-failing-combat-inauthentic-behaviour-online; and Rolf Fredheim, "Robotrolling 1/2019," NATO Strategic Communications Centre of Excellence, 2019, https://www.stratcomcoe.org/robotrolling-20191.

52 This idea may be credited to Jonathan Herrmann of George Mason University.

CARNEGIE
ENDOWMENT FOR
INTERNATIONAL PEACE

1779 Massachusetts Avenue NW | Washington, DC 20036 | P: + 1 202 483 7600

CarnegieEndowment.org