CARNEGIE
ENDOWMENT FOR
INTERNATIONAL PEACE

WORKING PAPER

SEPTEMBER 2020

Future Threats, Future Solutions | #3

The EU's Role in Fighting Disinformation: Developing Policy Interventions for the 2020s

James Pamment

The EU's Role in Fighting Disinformation: Developing Policy Interventions for the 2020s

James Pamment

Carnegie does not take institutional positions on public policy issues; the views represented herein are those of the author(s) and do not necessarily reflect the views of Carnegie, its staff, or its trustees.

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 Project

This paper is the first of a three-part series called Future Threats, Future Solutions that looks into the future of the European Union's (EU) disinformation policy.

This series was commissioned by the European External Action Service's (EEAS) Strategic Communications Division and prepared independently by James Pamment of the Partnership for Countering Influence Operations (PCIO) at the Carnegie Endowment for International Peace. Over one hundred experts, practitioners, and scholars participated in five days of workshops, made written submissions, and/or completed surveys that fed into these papers. The resulting publications are the sole responsibility of the author and do not reflect the position of the EEAS or any individual workshop participant.

The first paper, "Taking Back the Initiative," focuses on future threats and the extent to which current EU disinformation policy instruments can meet the challenge. With the coronavirus pandemic erupting during the drafting of these papers, the overview of current instruments has been supplemented with discussion of lessons learned from the ongoing experience of this crisis. This first paper also outlines the overall policy recommendations detailed in the three papers.

The second paper, "Crafting an EU Disinformation Framework," establishes terminology and a framework around which EU institutions can organize their disinformation policy. The paper begins with a discussion of terminology and then outlines the ABCDE (actors, behavior, content, degree, effect) framework for analyzing influence operations. This supports further analysis of areas of institutional responsibility, including ownership of different aspects of the disinformation policy area.

The third paper, "Developing Policy Interventions for the 2020s," outlines three areas of intervention necessary for developing an EU disinformation policy capable of meeting future threats. The first is work that deters actors from producing and distributing disinformation. The second consists of nonregulatory interventions, which focus primarily on policies that can be enacted informally with stakeholders. The third covers regulatory interventions, including legislative responses based upon an auditing regime.

Introduction

The European Union (EU) needs a well-conceived and forward-looking policy for countering threats in the information space, especially those posed by disinformation, influence operations, and foreign interference. Because of the amorphous and ever-changing nature of the threat, EU officials and their counterparts in EU member states would do well to craft an approach that draws on a variety of effective tools, including strategies for altering adversaries' behavior, nonregulatory principles and norms to foster a well-functioning digital public sphere, and regulatory interventions when necessary to ensure that digital platforms uphold suitable norms, principles, and best practices.

Influencing Adversaries' Calculus

As a rule, the EU does not practice deterrence. Furthermore, the applications of traditional deterrence theory to the challenge of disinformation are limited.[1] The EU is, however, capable of working in ways that can collectively influence adversaries' calculus to dissuade them from spreading disinformation, running influence operations, or conducting foreign interference. Ultimately, a significant portion of disinformation policy should aim to achieve this effect. The activities discussed in this paper sketch out an overview of options, with the aim of eventually crafting a framework of cumulative deterrence.[2] The EU's policy on disinformation should build on this clear aim of deterring adversaries, using all available policy instruments.[3]

The benefit of the approach outlined here is that misinformation and disinformation are treated primarily as problems of democracy to be dealt with by improving the health of public debate in EU member states. Influence operations and foreign interference are treated as security concerns in the context of attempting to influence the calculus of adversary actors. This latter approach acknowledges that actor-specific knowledge and countermeasures are a necessary foundation of disinformation policy.

The EU should consider its interventions from the perspective of raising costs and denying benefits to adversary actors to protect the integrity of public debate and other key national interests. Interventions should be designed to influence the calculus of adversary actors so they no longer perceive disinformation, influence operations, and manipulative foreign interference campaigns as beneficial courses of action.

Some current EU bodies, such as the Strategic Communications Division and the East StratCom Task Force of the European External Action Service (EEAS), make strong contributions to developing societal and institutional resilience. Modern, data-driven strategic communication and public diplomacy are central to maintaining and projecting these capabilities. To further develop their potential, EU policymakers should reconsider how these elements contribute to a cumulative posture aimed at dissuading adversary actors from spreading disinformation and conducting influence operations and foreign interference. They should consider regulatory and nonregulatory interventions related to disinformation not merely in terms of data transparency but in terms of geopolitics. In short, raising costs, denying benefits, and denying capabilities should be core motivations driving policy interventions.

Actor-specific policies for disinformation are required in certain instances. EU policy toward Russia and other identified adversary actors should be assertive in integrating disinformation-related concerns into the EU's broader engagement posture, including the use of sanctions. Lines in the sand should be established at the political level to demonstrate the resolve of member states around, for example, election interference. Integration with intelligence, hybrid, and cyber capabilities—which is not discussed in detail here—should also be considered.

Resilience

The EU can take steps to influence a given adversary's calculus by demonstrating preparedness, capabilities, and resilience. This means that the societies of EU member states are prepared for interference and can withstand, or bounce back from, whatever unexpected attacks an adversary designs. Resilience therefore refers to both the quality of contingency planning (risk mitigation) and the ability to adapt to and recover from successful attacks (assertiveness, agility, and resolve). Elements of fostering such resilience include communicating preparedness, building capacity, assessing threats, analyzing adversaries' influence networks, and communicating with adversaries.

- **Communicating preparedness** involves assertive acts of communication designed to mitigate risk. Achieving such preparedness depends on professional and credible communications services and the ability to assert and project one's story and identity through strategic narratives, public diplomacy, and branding. This critical task also involves generating public awareness of disinformation through credible and trusted sources, including basic educational steps such as media literacy and online hygiene. Assurance also plays a role on this front by demonstrating commitment, resolve, and capability to domestic, allied, and adversary publics. The European Commission's Directorate-General for Communication (DG COMM) and the Directorate-General for Education, Youth, Sport and Culture (DG EAC), as well as the EEAS's Strategic Communications Division already conduct significant work in these areas, in conjunction with the East, South, and Western Balkans StratCom task forces.

- **Building capacity** entails making assessments of risk and contingencies, developing appropriate coordination capabilities, and establishing robust partnerships with international peers, the private sector, and civil society. This series of papers heavily emphasizes the importance of stakeholder management to future EU policy, particularly in improving the relationship between digital platforms and researchers. The European Commission's Directorate-General for Communications Networks, Content and Technology (DG CONNECT), the Directorate-General for European Neighbourhood Policy and Enlargement Negotiations (DG NEAR), the Joint Research Centre (JRC), and the EEAS all perform significant roles in this regard.

- **Assessing threats** in terms of their likelihood and severity is another crucial task. It involves developing situational awareness and observational thresholds through digital monitoring capabilities, for example. This means cultivating an understanding of adversaries' capabilities, intentions, and opportunities in the area of disinformation, and doing so relies upon cooperation between open-source and secret intelligence as a basis for speedy attribution and communication of threats. The East StratCom Task Force currently coordinates some open-source tasks with a specific focus on Russia. However, these capabilities could be further developed, as could actor-specific knowledge and strategies for other adversaries. The Rapid Alert System also provides a coordination function for EU member states, though it is not currently used to its full potential.

- **Analyzing adversaries' influence networks** focuses on primarily domestic and transnational groups that support a given adversary's goals directly or indirectly. Work can be done to analyze an adversary's proxies in one's own environment and to better understand influence networks in foreign states as potential tools of influence. Best practices derived from countering violent extremism, such as understanding the motivations and grievances of affected populations and supporting moderate voices within vulnerable communities, are examples of how public and cultural diplomacy can build societal resilience and help counter these influence networks. Fighting organized crime also offers examples of best practices when it comes to analyzing such networks. Media monitoring by the East StratCom Task Force has analyzed the networks through which pro-Kremlin narratives have spread, though this function was deprioritized in 2018.

- **Communicating with adversaries** involves making use of one's own formal and informal transnational influence networks—like those built up via diaspora links, diplomacy, business ties, student exchanges, and public diplomacy, for instance—to engage in a dialogue with a given adversary. Such interactions may be valuable in times of heightened tensions or crisis. The EEAS and DG NEAR currently have significant investments in public diplomacy campaigns in the eastern neighborhood, for example. However, these campaigns struggle to reach relevant audiences at present.

Resilience is about good governance and preparedness. This is a strength upon which EU institutions should build. Some current EU instruments already make strong contributions in this area. Modern, data-driven strategic communication and public diplomacy are central to maintaining and projecting these capabilities, and these assets should be bolstered and integrated with the tools of countering disinformation. To further develop their potential, EU policymakers should reconsider how these elements contribute to a cumulative posture aimed at dissuading adversary actors from spreading disinformation and conducting influence operations. In particular, they should reconsider resourcing levels with a view toward ensuring an appropriate balance between actor-agnostic and actor-specific risk assessments.

Incentives

Aside from building societal resilience, offering incentives is a means of encouraging desirable behavior without resorting to threats. Such inducements are designed to move different elements to an adversarial relationship beyond its usual persistent problems. At a most basic level, such incentives include payments and rewards. They can also involve working together on common problems where both parties stand to gain.

The EU has a considerable range of incentives that it can offer adversary actors, including painting compelling visions of the future, brainstorming common projects, offering inclusion in a community, designing suitable reward structures, and reciprocating de-escalation.

- **Painting compelling visions of the future**: One form of positive reinforcement involves using strategic narratives, traditional diplomacy, and public diplomacy to communicate a positive desired end state for an adversarial relationship. DG NEAR, the EEAS, and the StratCom task forces work extensively in this area, for example, in relation to the EU's neighborhood.

- **Brainstorming common projects**: Actors work together in areas that are uncontroversial and where both parties benefit. Such arrangements ensure that points of contact are maintained even during tense moments. This work is often conducted under the umbrella of public and cultural relations, as well as scientific and educational cooperation, for example. Again, this is an area where support to the EU delegations, particularly in the EU's neighborhood, provides positive incentives for maintaining a good relationship with Brussels and EU member states.

- **Offering inclusion in a community**: A clear pathway to joining an international community or agreement can encourage an adversary to reduce its disruptive activities. Such a community or agreement builds upon a vision of a prosperous future shared by both parties. EU accession dialogues, for example, provide public diplomacy opportunities in this vein.

- **Designing reward structures**: The EU can reward actors for good behavior. Rewards may include market access, aid, foreign direct investment, technical assistance, or other similar measures. The EU has several inducements it can offer through development assistance and cultural diplomacy to connect with otherwise hard-to-reach grassroots audiences.

- **Reciprocating deescalation**: The EU can reciprocate an adversary's willingness to reduce its disruptive activities by reducing the union's own disruptive activities, as defined by the adversary (a way of saying that if you stop doing this, we'll stop doing that). The Kremlin has, for example, repeatedly complained about the East StratCom Task Force, and a clear response to these complaints should emphasize that when the Kremlin stops spreading disinformation, the task force will no longer be necessary.

An incentives-based approach to deterring disinformation and influence operations requires a challenging amount of coordination between EU member states and EU institutions. The easiest wins to be had are in development assistance, public diplomacy, cultural diplomacy, and strategic communication. The premise of such incentives does, however, assume an ability to create attractive and realistic narratives and projects that engage parts of external populations that might otherwise be tempted to engage in disinformation activities, as well as state-level activities. In short, this approach is desirable for many reasons, and it is strongly tied to the EU's external image.

Denial of Benefits

Disinformation is a low-risk, high-reward endeavor. Raising the costs of conducting certain malign activities can remove or at least lower incentives for adversaries to employ them. The most realistic tools for imposing such costs include exposing disinformation actors, as the East StratCom Task Force does to a certain extent, and taking additional steps together with digital platforms to raise the costs of manipulating digital media. Several supporting regulatory and nonregulatory recommendations are presented here. These recommendations also involve communicating and demonstrating to adversaries that they will not reap the benefits they desire from pursuing these activities. Key elements of denying adversaries the benefits of conducting disinformation campaigns and influence operations include projecting solidarity and resolve, framing disinformation campaigns, pursuing technical attribution, and considering political attribution.

- **Projecting solidarity and resolve**: Demonstrating a high-cost response from multiple actors should be the EU's main strength. On matters like election interference, the EU should clearly draw lines in the sand to signal to other countries what behavior it deems unacceptable and how it would respond to deter such behavior. This practice should particularly emphasize solidarity and resolve around interference targeting individual member states.

- **Framing disinformation campaigns**: This task refers to strategic communication that shapes audiences' interpretations of actors and events in ways that support overall efforts to isolate, undermine, or damage EU adversaries. The European Commission's DG COMM and the EEAS's Strategic Communications Division have key roles to play on this front.

- **Pursuing technical attribution**: Forensic evidence of an influence operation that links specific activities to a given adversary can help raise the costs of conducting such operations. Revealing the actors who engage in such conduct and the techniques they employ can help prevent such malign actors from wielding such capabilities and methods in the future, and such revelations also can provide the evidence to support political attribution. Technical attribution can also augment deterrence by informing the adversary that its activities can be carefully traced and, presumably, that its network can be targeted by way of a denial-based response. Technical attributions do not have to be made public but can be shared among trusted stakeholder groups. The EU can collaborate with members states, researchers, and civil society organizations with open source intelligence capabilities to support public debate around the attribution of threat actors.

- **Considering political attribution**: A public accusation supported either by rhetoric or by technical evidence involves censuring an adversary to damage its reputation. The East StratCom Task Force performs this function for the EU, albeit with a disclaimer that it does not represent EU policy. As set out in the first paper in this series, "Taking Back the Initiative," political processes for exposing state actors beyond Russia must improve.

Denial of Capabilities

Preventing an adversary from using its capabilities involves communicating and demonstrating that the EU will respond to attacks in ways that disable or degrade an adversary's current means of spreading disinformation. This task would entail, for example, working with digital platforms to close vulnerabilities that enable the spread of disinformation at scale. This objective is a focal point of the regulatory and nonregulatory interventions outlined in this paper. Specific elements of these efforts include upholding rights and responsibilities, coordinating on takedowns, demonetizing disinformation, and cultivating covert capabilities.

- **Upholding rights and responsibilities**: The EU can emphasize behavioral norms in democratic societies, in the international community, and on digital platforms. Some representatives and proxies of adversary countries tend to invoke their rights when acting in Western markets (citing freedom of speech, for instance) but then behave in an irresponsible manner (by spreading disinformation, for example). Nonregulatory efforts to strengthen digital platforms' terms of service in line with fundamental freedoms could support efforts to deplatform actors who exploit digital services.

- **Coordinating on takedowns**: Takedowns refer to the coordinated removal of online content or the banning of social media accounts often in collaboration with the hosting digital platforms. The EU's main capabilities here would be in building robust channels of communication and shared standards with digital platforms. In particular, the EEAS should seek a much closer relationship with industry actors.

- **Demonetizing disinformation**: Aspects of digital platform manipulation are motivated by economic incentives. Denying online actors the ability to generate income from disinformation should be a priority for EU policymakers.

- **Cultivating covert capabilities**: Covert denial capabilities should be integrated into the EU's overall posture for specific threat scenarios.

Denial by Punishment

The EU's policy toward Russia and other identified adversary actors should be assertive in integrating disinformation-related concerns into the union's broader engagement posture. Some of these efforts can only be achieved at the highest political levels. However, many regulatory and nonregulatory interventions would be appropriate not merely from a data transparency perspective but also from a geopolitical perspective. Raising costs, denying benefits, and denying capabilities should be among the core motivations driving EU policy interventions, but these tactics must ultimately be backed up by actor-specific punitive measures for them to be credible.

- **Enacting necessary expulsions and travel bans**: Usually associated with diplomats, suspected terrorists, or official representatives of organizations closely associated with an adversary, expulsions and travel bans deny individuals freedom of movement and could be applied to the originators of a disinformation campaign.

- **Levying sanctions**: International sanctions can be applied in areas such as diplomacy, trade, and even sports to affect a targeted adversary's reputation and restrict its access to markets. Sanctions against individuals can be directed against high-ranking officials or supporters of an adversary, restricting their movement and access to markets. Enacting such sanctions would require a political decision.

- **Prosecuting violators**: Such prosecution would entail formally charging or prosecuting individuals or organizations via independent legal processes for illegal behavior related to, for example, spreading harmful online content or sedition.

Actor-Agnostic Versus Actor-Specific Strategies

Some of the activities outlined above aimed at deterring malign actors from spreading disinformation and engaging in foreign interference are actor-agnostic—that is to say, they apply to all adversaries. Many commercially driven actors, including some state actors, can be targeted with the same measures. In other cases, however, actor-specific strategies are needed, since the broader considerations of foreign policy and geopolitics are highly salient to deterrence work.

Disinformation policy should be premised on considerations of which elements of this toolkit can be used to develop a posture capable of altering the calculus of a given adversary and under which circumstances. Political decisions about lines in the sand would be desirable. The objective is not simply to accept the disinformation activities and influence operations of adversaries but to develop a strategy to change their behavior over time by pushing back on certain patterns. The European Centre of Excellence for Countering Hybrid Threats recently shared the "4S model" with EU member states.[4] EU member states and institutions could enact this model and integrate its steps into their broader efforts to counter disinformation.

Nonregulatory Interventions

Several types of interventions to counter disinformation and influence operations do not require regulation and would arguably serve the interests of the EU and tech platforms far more effectively than current structures. For example, in the voluntary EU Code of Practice on Disinformation, signatories reported their data according to their own definitions of disinformation and without independent verification. This is also a common practice in general reporting on platform transparency. Feedback from the process of formulating the code of practice indicates that stakeholders would benefit from common definitions and some form of verification of self-reported data.[5]
One simple but significant nonregulatory intervention would be for the EU to establish a common terminology for describing and capturing metrics related to disinformation, as laid out in the second

paper in this series, "Crafting an EU Disinformation Framework." A second necessary step is to define in what ways self-reported data should be verified. Such verification should include proposals for reporting formats and standards so as to form a basis for collecting consistent evidence that can be used to inform policymaking.[6]

For many of these nonregulatory steps to be realistic, a new collaborative process designed to replace and build upon the code of practice is required. Platforms argued that EU policymakers did not always ask the right questions when they were crafting the code of practice, which made it impossible for them to select the most relevant data. A more collaborative, iterative process is needed to ensure that the EU, its member states, and digital platforms can fully understand and adapt to one another's needs and present limitations. In particular, stakeholders need to achieve a better understanding of data collection, policy processes, and product development at platforms if governments and platforms are to find a practical means of attaining the common goal of countering disinformation. Any long-term vision of EU policy on disinformation should be backed up by a consultative process with platforms that involves recurring check-ins and opportunities for iterative adaptations when necessary.

Refining Platforms' Terms of Service

The terms of service and community standards of digital platforms provide one form of nonregulatory intervention. Platforms define acceptable use of their services to establish enforceable norms. Breaching these norms can, for example, lead to the suspension or closure of accounts or pages and the deletion of norm-offending posts. It is up to each platform to define acceptable norms and usage. These standards can vary widely, as the major platforms have very different standards and enforce different norms based, in many cases, on user expectations of what is appropriate or not on those platforms. The major platforms' different approaches to handling political advertising is one case in point.

When international law provides limited opportunities for responses to counter disinformation and influence operations, digital platforms can and do enforce policies that impact their spread. It may be desirable for the EU to take the lead in defining general guidance derived from appropriate legal and normative principles to support the development of clear definitions of harm and/or interference related to disinformation. Clearly, such efforts should respect differences among platforms and the expectations that users have on them; hence, such guidance should consist of nonspecific, adaptable building blocks. It should include elements such as content moderation policies, enforcement priorities, and appeals procedures. In particular, this approach can be valuable in terms of shaping the behavior of emerging and future digital platforms as they grow. Such an initiative could help to strengthen platforms' resolve to enforce their terms of service.

Differentiating Between Various Groups of Platform Users

Audience size and influence could, and perhaps should, shape approaches to nonregulatory intervention when it comes to digital platforms' terms of service and community standards. Platforms are malleable and designed to be used in different ways by commercial and noncommercial users. Stakeholders should consider alternative methods of categorizing users based on their identities and behavior. For example, a user with 200 followers clearly has a different social role and set of responsibilities than a user with 200,000 followers.

Some people use the platforms to stay in touch with friends and family, while others essentially use them as infrastructure for publishing news or opinions. Methods of enforcing different terms of service, based on the identity, reach, and behavior of a given user rather than the platform they use, could be conducive to a more nuanced approach to how influence can be exerted on digital platforms.[7] The EU could develop general guidelines to support digital companies' adaptation of these principles in a manner appropriate to their platforms.

Promoting and Demoting Content

The algorithms that govern user experience on digital platforms are commercially sensitive intellectual property. Governments and researchers would like to understand them better, but platforms are naturally reluctant to reveal such information unless legally required to do so. It is widely understood that platforms can adapt their algorithms to promote and/or demote content based on assessments of whether it contains misinformation and disinformation. Such an approach does not infringe on fundamental democratic freedoms and allows the platforms to support a healthy form of public discourse. A transparent means of demonstrating the governing policies, implementation, and verification of these activities is essential to establishing an evidence base for determining the effectiveness of this approach.[8]

The EU could provide valuable guidance on what constitutes healthy public discourse, which platforms could carry forward in the development of this crucial method of intervention. This principle should apply equally to human-led and AI-filtered content moderation. Platforms should consult independent experts on their content moderation policies and be bound to share with outside researchers and auditors the borderline cases they escalate internally.

Conducting Takedowns

Digital platforms conduct takedowns in a variety of ways. Usually, they either identify or receive a tip about a potential case. Threat investigation teams then assess the case and, in a coordinated manner, remove all associated users, pages, and/or content. Some platforms release the depersonalized data to the research community, while others restrict it to a select group of researchers. Several big-picture questions remain. These questions include:

- What content and behavior motivates a takedown in light of fundamental freedoms?

- What is the threshold for a takedown as opposed to other noninvasive interventions such as content demotion?

- How can the overall processes and framework for adjudicating takedown cases, including relevant appeals procedures, be defined and independently assessed as appropriate and fit for purpose?

- How should authentic users caught up in a case be notified?

- How and at what stage should relevant governments be informed?

- How should lessons learned be shared with product teams and how should such lessons be evaluated?

Engaging in Evidence-Based Attribution

The process of attributing actors that produce and/or disseminate disinformation is highly sensitive due to technical, political, and legal concerns. Currently, some EU member states attribute adversary actors in certain instances while others do not. Within EU institutions, the East StratCom Task Force explicitly attributes pro-Kremlin media, though with a disclaimer that this designation does not represent official EU policy. In a recent address, European Commission Vice President for Values and Transparency Věra Jourová explicitly pointed to Russian and Chinese disinformation in Europe.[9] Digital platforms all have their own differing processes for attribution. More work should be done to clarify and where possible harmonize the meanings of attributions to better inform the public about the sources of disinformation. In particular, a set of clear ground rules and guiding principles would be valuable for setting basic standards for the types of information shared with the public—and, potentially, through more secure means—within the stakeholder community.

Attribution involving disinformation should be considered in three layers: information distilled from open sources, proprietary sources, and secret (classified) sources.

- **Open sources**: Often nongovernmental organizations, journalists, and researchers use techniques drawn from investigative journalism combined with digital skills to reveal hidden information. Currently, EU public attributions, including those by the East StratCom Task Force, rely primarily on this approach.

- **Proprietary sources**: Often digital platforms and private intelligence companies use their own data combined with business and commercial intelligence sources to determine patterns of behavior, often with a focus on the back end and underlying infrastructure of an influence operation. EU collaboration in this area could be improved so that such efforts can support and amplify the attributions of others and so that EU stakeholders can compare these findings with its own.

- **Secret sources**: Primarily governments and military personnel use classified signals intelligence and cyber-related capabilities to monitor and intervene in cat-and-mouse-style interactions with adversaries. The EU should aim to support and amplify the attributions of others where appropriate.

Collaborating on Research

The EU should develop guidelines for collaborative models of cross-sector research to support high-quality analysis on disinformation and influence operations. Any attempts to foster such collaboration should distinguish between the needs of the operational research community, which typically produces briefings on short time frames, and the academic research community, which produces peer-reviewed academic research often over a period of years.

According to a recent study, these collaborative research models can rely on formal or informal institutional arrangements that have been achieved productively in U.S. defense and intelligence circles. Collaborative models face the challenge of allowing researchers to maintain their credibility and independence while engaging with industry or government actors, ensuring data security and overcoming structural barriers that may impede successful collaboration.[10]

Advocating Best Practices for Political Parties and Politicians

A great deal of confusion and controversy arises over the fact that governments, politicians, and political parties can easily be accused of spreading disinformation. At times, these accusations are justified. While more formal guidance is the responsibility of each EU member state, the EU itself could support this process for member states and other third parties by developing best practices so that political parties do not spread misinformation or disinformation or mimic the illegitimate techniques of influence operations. Political advertising is a key dimension of this area of work. Such efforts would also help platforms develop and implement their own content policies related to government figures and political parties.

Addressing Exceptional Circumstances

Events like European Parliament elections or the crisis surrounding the coronavirus pandemic should be treated as exceptional circumstances in which the EU and digital platforms engage in an intensified form of collaboration. This collaboration already exists to a certain degree but could be developed particularly for crises like the coronavirus pandemic. Examples of forms such collaboration could take include:

- Fast-tracking the involvement of EU-based researchers in takedowns,
- Organizing biweekly information-sharing briefings, including on leads/ongoing trends (on takedowns and other relevant topics),
- Reporting on overt state media campaigns (including their tactics, reach, and the potential use of fake engagement or spamming tactics), and
- Considering enhanced state media policies, such as temporary fact checking and labeling in conjunction with other stakeholders.

Regulatory Interventions

The nonregulatory measures listed above are examples of the kinds of solutions that collaborative, good-faith, and inclusive approaches to the problem of disinformation can offer. In some cases, including in these nonregulatory examples, policymakers may desire a harder line of regulation to ensure cooperation.

However, the exact model of regulatory intervention most appropriate to mitigating the effects of disinformation remains unclear. Established models drawn from media regulations, financial regulations, or cybersecurity measures could provide a guide. Yet disinformation may ultimately prove to be a unique problem warranting unique solutions.

The following discussion will focus not so much on details as it will on general principles that could help inform a suitable regulatory stance. The basic process described here would rely on three main steps. First, the EU (or other regulator in question) would set out shared expectations, key performance indicators, and/or harm mitigation requirements to digital platforms. Second, the digital platforms would report at regular intervals on their compliance with these expectations and requirements. And third, a control or auditing mechanism would be established to independently verify such compliance. (The exact methods of compliance and verification are not discussed further in this paper as this is essentially a technical discussion.)

Assuming an appropriate method is found, the following areas of regulation should be considered: independent oversight, capability transparency, duty of care, data transparency, social media manipulation, and a definition of breaches.

Independent Oversight

An independent body should be responsible for overseeing and implementing the EU's regulatory regime. It may also be appropriate for this body to oversee and implement nonregulatory interventions so as to ensure that relevant norms and principles are applied consistently and to provide a single point of contact and expertise for external stakeholders. This body should work with researchers to establish a progressive research agenda designed to analyze, verify, and predict trends in disinformation. Its primary objective should be to develop an evidence base to inform policymaking; auditing and data access are a means to that end, but not an end in themselves. An existing body such as the European Regulators Group for Audiovisual Media Services could be tasked with performing that mediating function between digital platforms and EU member states on matters of data compliance.

Capability Transparency

The appropriateness of internal platforms' processes for managing the risks they deal with presents a major question for EU oversight. In particular, it would be desirable for policymakers to have insight into how platforms manage risk, including on questions such as staffing and resourcing levels; risk assessments; and procedures for identifying, analyzing, and removing disinformation.[11]

Duty of Care

An associated avenue for the EU to develop regulatory interventions is with respect to digital platforms' duty of care. This term refers to the legal responsibility of a person or organization to avoid any behaviors or omissions that could reasonably be foreseen to cause harm to others. An auditing regime could be established on the basis of a graduated list of harms similar to, for example, those defined in the United Kingdom's "White Paper on Online Harms."[12] This list of defined harms should be based on breaches of fundamental rights to expression, privacy, and political participation. Policymakers could then hold platforms accountable for demonstrating their efforts to protect users from breaches of these rights in line with a duty of care.

Data Transparency

Access to digital platforms' proprietary data is one of the main points of tension in the relationship between these platforms and governments, civil society, and independent researchers. Platforms frequently refer to privacy concerns, including the EU's General Data Protection Regulation (GDPR), as hindrances to their ability to share data. Researchers and EU officials disagree that the GDPR hinders access to platform data for research purposes. It should be within the EU's power to clarify this point of contention and if necessary make amendments to the GDPR to overcome any such obstacles.

Initiatives involving partnerships between universities and other stakeholders, such as Social Science One at Harvard University, have encountered serious technical and governance challenges, particularly in relation to privacy concerns.[13] A principle of transparent and equitable access for researchers should be a cornerstone of future policy, though the exact format this principle should take could emphasize access to *information* rather than *data* per se.[14] For example, France has tested an online front end that enables searches of platform data and gives aggregated results without granting access to the raw data itself. If such a format were considered desirable, access to the back end could still be granted to EU officials and/or member-state policymakers.

Areas of data transparency that should be considered include sample data, advertising transparency, and aggregate platform information.

- **Sample data**: Researchers and auditors should have access to samples of reported, suspended, removed, and restored online content to facilitate research and better scrutinize platforms' content policies.

- **Advertising transparency**: Auditors and researchers should also enjoy access to paid advertisements that appear on digital platforms, along with data about a given ad's purchaser, intended audience, and actual audience.

- **Transparency on influence operations**: Researchers and auditors ought to have access to information about post, page, and account takedowns, user reports, and platform actions under policies related to influence operations and foreign interference.

Social Media Manipulation

One of the areas in severe need of regulatory intervention is market exchanges involving social media manipulation, that is, the legal purchasing of fake engagements by companies that specialize in manipulating digital platforms.[15] EU regulation in this area could seek to prohibit the selling and/or purchasing of social media manipulation services, to prohibit the generation of income from the selling or purchasing of such services, and to require digital platforms to report known providers of social media manipulation services to legal authorities. Regulations should also disincentivize the spread of disinformation by adversary actors with economic incentives by intervening to disrupt the monetization of disinformation websites, especially through online advertising, for instance.

Defining Breaches

A harder regulatory approach requires that the EU define breaches (particularly regarding privacy and protection from harm) and how such breaches should be resolved. Punitive measures may be necessary. However, a broader question is how a duty of cooperation among platforms and the EU and its member states can be established. The best means of mitigating the societal impact of disinformation will always be through a broad stakeholder community working together with a shared vision. A focus on improvement, learning, and iterative development of definitions, interventions, and other relevant measures should be the goal of these relationships. Defining and censuring platform responsibilities in the case of breaches is therefore a particularly challenging aspect of developing regulatory interventions.

Conclusion

The challenges facing the EU in developing a new disinformation policy are substantial but should be informed by a simple question. If not the EU, then who? Which other international actor is currently capable of setting the agenda, establishing norms, and enforcing regulations to protect democracy from those who would diminish it? Done correctly, this policy may become a replicable model for many countries outside of the EU looking for effective, realistic, and balanced ways to counter online disinformation.

About the Author

James Pamment is a nonresident scholar in the Technology and International Affairs Program at the Carnegie Endowment for International Peace and co-director of the Partnership for Countering Influence Operations there.

Notes

1 Vytautas Keršanksas, "Deterrence: Proposing a More Strategic Approach to Countering Hybrid Threats," European Centre of Excellence for Countering Hybrid Threats, March 2020, https://www.hybridcoe.fi/wp-content/uploads/2020/03/Deterrence.pdf; Mikael Wigell, "Democratic Deterrence: How to Dissuade Hybrid Interference," Finnish Institute of International Affairs, September 2019, https://www.fiia.fi/wp-content/uploads/2019/09/wp110_democratic-deterrence.pdf; and James Pamment and Henrik Agardh-Twetman, "Can There Be a Deterrent Strategy for Influence Operations?" *Journal of Information Warfare* 18, no. 3 (Winter 2019): 123–135, https://www.jinfowar.com/journal/volume-18-issue-3/can-there-be-deterrent-strategy-influence-operations.

2 Doron Almog, "Cumulative Deterrence and the War on Terrorism," *Parameters: United States Army War College Quarterly* 34, no. 4 (Winter 2004–2005): 4–19, https://www.hsdl.org/?abstract&did=453973; and Uri Tor, "'Cumulative Deterrence' as a New Paradigm for Cyber Deterrence," *Journal of Strategic Studies* 40, no. 1 (January 2017): 92–117, https://doi.org/10.1080/01402390.2015.1115975.

3 This section borrows heavily from Pamment and Agardh-Twetman, "Can There Be a Deterrence Strategy for Influence Operations?"

4 (The document referenced here is not publicly available.) European Centre of Excellence for Countering Hybrid Threats and the Community of Interest of Hybrid Influencing, "Deterring Hybrid Threats: A Playbook for Practitioners" (Helsinki: European Centre of Excellence for Countering Hybrid Threats, 2020).

5 James Pamment, "EU Code of Practice on Disinformation: Briefing Note for the New European Commission," Carnegie Endowment for International Peace, March 3, 2020, https://carnegieendowment.org/2020/03/03/eu-code-of-practice-on-disinformation-briefing-note-for-new-european-commission-pub-81187.

6 Mark MacCarthy, "Transparency Requirements for Digital Social Media Platforms: Recommendations for Policy Makers and Industry," Transatlantic Working Group on Content Moderation Online and Freedom of Expression, February 12, 2020, 16, https://www.ivir.nl/publicaties/download/Transparency_MacCarthy_Feb_2020.pdf.

7 Jonathan Hermann, "With Greater Audience Comes Greater Responsibility," Carnegie Endowment for International Peace, April 23, 2020, https://carnegieendowment.org/2020/04/23/with-greater-audience-comes-greater-responsibility-pub-81582.

8 MacCarthy, "Transparency Requirements for Digital Social Media Platforms," 8, 15–16.

9 Věra Jourová, "Opening Speech of Vice-President Věra Jourová at the Conference 'Disinfo Horizon: Responding to Future Threats,'" European Commission, January 30, 2020, https://ec.europa.eu/commission/presscorner/detail/en/speech_20_160.

10 Jacob Shapiro, Michelle P. Nedashkovskaya, and Jan G. Oledan, "Institutional Models for Understanding Influence Operations: Lessons from Defense Research," Carnegie Endowment for International Peace, June 25, 2020, https://carnegieendowment.org/2020/06/25/collaborative-models-for-understanding-influence-operations-lessons-from-defense-research-pub-82150.

11 Ranking Digital Rights has developed a methodology for comparing some of these issues. See "Recommendations for Government and Policymakers," Ranking Digital Rights, https://rankingdigitalrights.org/governments-policy.

12 UK Department for Digital, Culture, Media and Sport and Home Office, "Online Harms White Paper," updated February 12, 2020, https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper.

13 Harvard University Social Science One, “Our Facebook Partnership,” https://socialscience.one/our-facebook-partnership.

14 Ben Nimmo, “Investigative Standards for Analyzing Information Operations,” draft presented at the Brookings High-Level Transatlantic Working Group on Disinformation and Emerging Technology, February 19–21, 2020, forthcoming.

15 NATO StratCom Centre of Excellence, “The Black Market for Social Media Manipulation,” NATO StratCom Centre of Excellence, November 2018, https://www.stratcomcoe.org/black-market-social-media-manipulation.

CARNEGIE
ENDOWMENT FOR
INTERNATIONAL PEACE

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

CarnegieEndowment.org