



RESEARCH  
ARTICLE



OPEN  
ACCESS



PEER  
REVIEWED

# Combating misinformation online: re-imagining social media for policy-making

**Eleni A. Kyza** *Cyprus University of Technology* Eleni.Kyza@cut.ac.cy

**Christiana Varda** *Cyprus University of Technology*

**Dionysis Panos** *Cyprus University of Technology*

**Melina Karageorgiou** *Cyprus University of Technology*

**Nadejda Komendantova** *International Institute for Applied Systems Analysis*

**Serena Coppolino Perfumi** *Stockholm University*

**Syed Iftikhar Husain Shah** *International Hellenic University*

**Akram Sadat Hosseini** *University of Stuttgart*

**DOI:** <https://doi.org/10.14763/2020.4.1514>

**Published:** 21 October 2020

**Received:** 10 June 2020 **Accepted:** 7 August 2020

**Funding:** Co-Creating Misinformation-Resilient Societies (Co-Inform). Funded by the European Commission (grant agreement 770302).

**Competing Interests:** The author has declared that no competing interests exist that have influenced the text.

**Licence:** This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License (Germany) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. <https://creativecommons.org/licenses/by/3.0/de/deed.en>  
Copyright remains with the author(s).

**Citation:** Kyza, E. A. & Varda, C. & Panos, D. & Karageorgiou, M. & Komendantova, N. & Perfumi, S. C. & Shah, S. I. H. & Hosseini, A. S. (2020). Combating misinformation online: re-imagining social media for policy-making. *Internet Policy Review*, 9(4). DOI: 10.14763/2020.4.1514

**Keywords:** Misinformation, Policy-making, Internet policy, Social media

**Abstract:** Social media have created communication channels between citizens and policymakers but are also susceptible to rampant misinformation. This new context demands new social media policies that can aid policymakers in making evidence-based decisions for combating misinformation online. This paper reports on data collected from policymakers in Austria, Greece, and Sweden, using focus groups and in-depth interviews. Analyses provide insights into challenges and identify four important themes for supporting policy-making for combating misinformation: a) creating a trusted network of experts and collaborators, b) facilitating the validation of online information, c) providing access to visualisations of data at different levels of granularity, and d) increasing the transparency and explainability of flagged misinformative content. These recommendations have implications for rethinking how revised social media policies can contribute to evidence-based decision-making.

This paper is part of **Trust in the system**, a special issue of *Internet Policy Review* guest-edited by Péter Mezei and Andreea Verteş-Olteanu.

## 1. Introduction

Policymakers around the world are tasked with the responsibility of developing roadmaps to support democratic societies' decision-making processes but are also expected to guide reactions to high-risk situations in a rational, evidence-based approach. Social media are undisputable channels of communication, as evidenced by the huge numbers of everyday users, and their patterns of use (Perrin & Kumar, 2019; European Commission, 2018). In periods of heated public discussions, such as during major social crises, social media can be mined to understand trends, finetune policies, monitor the spread of misinformation, and guide actions at the collective level. Yet, in an era when opinions are formed and transformed online, policymakers seem to primarily use media to merely inform citizens, despite recommendations that they should be using social media to understand, connect and engage with citizens (Kapp, Hensel, & Schnoring, 2015). Avery (2017) surveyed 226 public information officers (PIOs) in the United States to understand how they use social media during public health crises; results indicate low monitoring of social media and point out that increased social media monitoring led to increased satisfaction with how these PIOs handled the crisis.

An important area of policy-making is the development of strategies to address the spread of misinformation surrounding important societal issues in several diverse contexts, such as politics, health, the environment, commerce, etc. However, the term misinformation has not been defined in a consistent manner in the literature (Vraga & Bode, 2020). Wardle and Derakhshan (2017) distinguish between three types of misrepresentation of information: mis-information, dis-information, and mal-information, which they define as follows:

- Mis-information is when false information is shared, but no harm is meant.
- Dis-information is when false information is knowingly shared to cause harm.
- Mal-information is when genuine information is shared to cause harm, often by moving information designed to stay private into the public sphere.

(Wardle & Derakhshan, 2017, p. 5)

Wardle and Derakhshan's definition of the misrepresentation of misinformation, and

its distinction from other types of what they call *information disorders*, focuses on the purpose of spreading false information and the intent to harm (or lack thereof). A similar definition has been shared in formal reports of the European Commission (High Level Group on fake news and online disinformation, 2018). Vraga and Bode (2020) discussed the contextualised and bounded nature of misinformation and offered a nuanced description of how to identify accurate or inaccurate information and address the changing context of misinformation, based on two main criteria: expert consensus and best evidence to-date. Vraga and Bode (2020), who also outline the challenges in characterising information as misinforming, explain that whether some information can be termed *misinformation* is dependent on the state of evidence, expert beliefs and the information environment in which misinformation occurs; thus, depending on the context the same information may be categorised in different ways.

Crises such as the COVID-19 pandemic further expose the problem of misinformation on social media. People around the world increasingly turned to the internet to actively seek information on pressing matters; at the same time, internet platforms, such as Facebook and YouTube, struggled to address the heightened demand for reliable information, as human moderation availability declined due to COVID-19 safety issues (Magalhães & Katzenbach, 2020). A recent review of the quality of online information on COVID-19 for lay people confirms prior reports on the abundance of low-quality, low-readability and low-reliability materials online, which can lead to the spread of misinformation, can increase panic among the general population, and may lead to non-rational actions (Cuan-Baltazar, Muñoz-Perez, Robledo-Vega, Pérez-Zepeda, & Soto-Vega, 2020). Public health is not, however, the only case where misinformation can spread with potentially disastrous effects. The mostly unregulated spread of misinformation on social media, the lack of formal gatekeepers, and the potential to influence major political events, such as Brexit and the US 2016 and 2020 elections, indicate that policymakers should be cognisant of what is happening online and should be ready to mine social media so that they can be in direct communication with the real world and reach informed decisions on how to address potentially problematic areas. At the same time, policymakers should be careful in the recommendations they make, so that the trust of the public in the information shared online, or in the institutions themselves, is not undermined, as happened in the case of the Chinese government censorship of early COVID-19 warnings by a young doctor. According to Larson (2020) policymakers should be careful in how to respond to misinformation, and especially the sanctions they may decide to impose, in order to maintain the people's trust, uphold democratic ideals, and avoid undue spread of misinformation

(such as mistakenly believing that drinking bleach will cure COVID-19) that might have potentially disastrous consequences.

Despite the vast amount of misinformation on online media, there are currently no clear policies or recommendations for determining how to manage misinformation on social media platforms, how to deal with it when detected, what legal frameworks and ethical issues to consider, how to disseminate corrective information, and how to encourage citizens to read or share corrective information as soon as it becomes available. Doing so is proving difficult, as social media platforms were not designed with policy-making in mind, having instead focused on promoting users' personal interactions with each other. In addition, case studies such as the one published by Jabbar, La Londe, Debray, Scott, and Lubienski (2014), who investigated the use of evidence by policymakers in response to hurricane Katrina, contribute evidence that policymakers' access to evidence in New Orleans during the Katrina crisis was brokered by others and that the policymakers themselves used anecdotes or personal experiences to guide their decisions. Such reports suggest a need for tools to support first-person inquiring into primary evidence in order to understand the dynamic spread of information on social media and its potential consequences.

Vosoughi, Roy, and Aral (2018) examined the spread of false information on Twitter from 2006 to 2017 and concluded that false information spread much faster than the corresponding true information and was diffused to more people, more deeply and more broadly. Furthermore, Vosoughi et al.'s analyses showed that people, more than bots, were the culprits in spreading such misinformation. Psychology and communication research report that simply presenting people with corrective information does not change their fundamental beliefs and opinions and may even reinforce them (Flynn, Nyhan, & Reifler, 2017; Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012); these researchers cite psychological mechanisms such as the *continued influence effect* and *worldview backfire effect* to explain the persistence of misinformation despite the efforts to debunk it. In a meta-review of 32 studies with more than 6,000 participants, Walter and Tukachinsky (2020) report that even though correcting misinformation might be effective if attention is paid to the design characteristics of the corrective information, it is not fully effective. Some strategies seem to hold potential in successfully addressing misinformation online and correcting misperceptions, such as providing timely corrections, providing coherent explanations rather than simple refutes, exposing related but disconfirming stories, and revealing the demographic similarity of the opposing group (Nyhan & Reifler, 2015; Walter & Tukachinsky, 2020). Now, more than ever, it

seems more evident that policy-making in the online world should be a shared endeavour, one that does not follow the conventional paths of policy-making and one that requires an understanding of the complex life cycle of the spread and regulation of misinformation online.

The work presented in this paper investigates policy-making-related challenges about tackling misinformation within the rapidly changing online environment, and the implications for human policymakers and for developing internet platform policies. We define internet platform policies as practices and rules for managing misinformation on online platforms. This paper is not about targeting specific content areas of misinformation from the perspective of policy-making; rather it is related to what policymakers perceive as challenges and as areas for improving the policy-making process in relation to the propagation of information on social media. Therefore, this paper asks what policymakers in three national contexts identify as challenges to combating misinformation on social media, and what suggestions they may have to support evidence-based policy-making and the creation of a misinformation-resilient environment. These issues are at the core of this empirical paper. To the best of our knowledge, there is little published information that reports on what policymakers, themselves, indicate would be useful to them to address online misinformation, thus also providing insights into the types of policies that could be implemented on social media platforms for this purpose. In addition, this work seeks to inform policy development and the socio-technical design of software to address misinformative content (automatically and/or with user intervention) in the context of the European Horizon 2020 project Co-Inform ([www.coinform.eu](http://www.coinform.eu)).

## **2. Policymakers, platform policies, and misinformation on social media**

Policy-making is a broad construct that covers all areas of social activity, such as economy, health, education, foreign affairs, etc. Policymakers can be civil servants, and may sometimes be elected to serve this role. Misinformation on social media can have far-reaching consequences and, as such, is of interest to policymakers. At the same time, inviting citizen participation in policy-making can enhance policymakers' understanding by offering diversity of ideas and new insights into complex problems from the stakeholders' point of view (Fischer, 2003). This new view of participatory policy-making can lead to increased engagement from the public and better-informed decision-making that reflects the needs of the people.

In this work we use the terms *policymakers* and *policy-making* to refer to human ac-

tors involved in formulating policies related to issues of societal importance, and the term *policies* to refer to the outcome of such governance decisions, online or offline. Internet platform policies seeking to mitigate the spread of misinformation on online media can take a variety of forms, which differ in the extent to which information is regulated, who is performing the regulation, and the role of the user in this process. Policies can be instituted via traditional channels, such as national governments, but are also reported to be increasingly automated, defined by the industry and internet platforms such as Facebook or Google (Picard & Pickard, 2017), with some researchers even discussing technology as a policymaker in its own right (Lessig, 1999; Just & Latzer, 2017). Braman (2016) explicitly identifies computer scientists and engineers as central policymakers on the internet, as their technical work, whether intended as such or not, creates the rules by which the internet works.

Policies for combating misinformation can be transparent or opaque. For instance, it is now well known that automated algorithms often decide who sees what on social media. While one of the reasons for the automatic selection and delivery of personalised news to users was to address information overload, this phenomenon is now widely discussed as potentially diminishing transparency and democratic participation on the internet, and as a policy-implementing, governance mechanism. Stray (2019) identified a taxonomy of six strategies that can contribute to combating disinformation, in particular, by examining and contrasting three cases: China as an example of an authoritarian government, Facebook as a global internet platform, and EU East StratCom Task Force, as a counter-propaganda group. The six strategies are *refutation*, *exposure of inauthenticity*, *alternative narratives*, *algorithmic filter manipulation*, *speech laws*, and *censorship*, all of which except censorship, as Stray argues, are legitimate approaches to countering disinformation. While in some countries with authoritarian regimes the internet is heavily regulated and even censored by the government, in Europe the EU's Regulation on open Internet access (Regulation (EU) 2015/2120) and net neutrality, requests that there are open channels of distribution and access of information, without any discrimination.

Recent collaborations between social media platforms and fact-checking organisations have led to a hybrid mode of regulation, in which human fact-checkers work with algorithmic outcomes to provide information to platforms, such as Facebook, on the veracity of news posts (Graves, 2018). Such efforts may lead to posts being removed from the social media platform. This work is also aligned with calls by the European Commission (High Level Group on fake news and online disinforma-

tion, 2018) for increased efforts to empower users to address misinformation and disinformation on social media. However, as contemporary policy-making involves active use of social media to monitor and engage with online discussions on important societal issues, understanding how social media regulation efforts can be extended to address the needs and challenges that policymakers face online is particularly important.

## 3. Methodology

### 3.1 Stakeholder input

This work is part of the interdisciplinary project “Co-Creating Misinformation-Resilient Societies” (Co-Inform), funded by the European Commission to develop online tools and policies that can support the civil society and professionals (policymakers and journalists) in mitigating the threat of misinformation on social media. Our work assumes that machine learning and human-aided approaches need to work side by side; it also aims to empower instead of censor the expression of ideas online. Our methods draw from stakeholder theory (Freeman, Harrison, Wicks, Parmar, & De Colle, 2010). To do so, ongoing work brings together three stakeholder groups (general citizens, journalists, policymakers from different fields) recruited in Austria, Greece, and Sweden, to co-create socio-technical solutions to misinformation on social media. In this paper, we primarily focus on insights from the policymakers’ stakeholder group, even though there are shared findings across groups which will be reported on elsewhere as our project progresses.

We view the involvement of stakeholders as a form of collaborative policy-making, for which there are compelling arguments and many initiatives at local and global levels (Innes & Booher, 2003). Such efforts seek to move away from positivist views of policy-making, which yield top-down regulations by governments or other organisations. Similarly, stakeholder theory, which was originally developed as a new paradigm to address rapidly changing business challenges, has been applied to several other issues of societal importance where understanding and addressing the needs of the target audience is becoming increasingly important (Freeman, Harrison, Parmar, & de Colle, 2010). This governance model involves co-creation processes which integrate views of various stakeholders and not only of “educated experts”. Frequently, outcomes of such decision-making processes include knowledge on the ground as well as enjoy higher levels of legitimacy and trust (Renn, 2008).

Innes and Booher (2003) argue that the complexity of current policy-making issues requires a multitude of contributions, due to a vast diversity of values, knowledge specialisation, needs, etc. By connecting to stakeholders, policymakers can more fluidly understand the dynamic landscape of interdependent interests, engage in constructive dialogue, and construct solutions that address real problems. Furthermore, stakeholder involvement in policy-making is aligned with the principle of affected interests (Fung, 2010) which states that democratic participation should allow individuals to take part in decision-making that pertains to their own interests.

### **3.2 Data collection**

For this study we collected the perspectives of citizens, journalists, and policymakers as concrete case studies at three co-creation sites in Austria (Vienna), Greece (Athens) and Sweden (Stockholm). Data were collected through focus groups and in-depth, personal interviews, amassing over nine hours of video data for focus groups and over 13 hours of videotaped interviews. The collected data were comparable across sites, as the same detailed protocols were provided to each of the co-creation site's workshop coordinators. Protocols focused on gathering data to gain insights into the contexts and situations that may lead to misinformation; the personal and professional practices in relation to credibility assessment of content found on social media; the types of evidence that are needed to convince a policymaker that a piece of information is trustworthy; the challenges faced when attempting to address misinformation online, and proposed solutions to this problem.

### **3.3 Participants**

Sixty-seven (67) representatives from three stakeholder groups (civil society, journalists, policymakers) participated in co-creation workshops at the three locations in an effort to understand practices, challenges and needs relating to misinformation-resilient behaviours on social media. While some activities included participants from all stakeholder groups, focus group discussions were conducted separately for each stakeholder group. Twenty-one (21) policymakers (Austria: 7; Greece: 9; Sweden: 5) participated in the three policymakers' focus groups, and eight policymakers (Austria: 2; Greece: 2; Sweden: 4) participated in in-depth interviews conducted after the conclusion of the workshops. Policymakers' backgrounds were diverse and included public sector officials as well as NGO representatives. At each of the sites we invited policymakers and practitioners, people who were affected by misinformation professionally and who felt committed to fight misinfor-

mation in their area. For example, in Austria policymakers felt committed to address the issues of misinformation which was relevant to the housing sector and migration. Their examples came from their own practice and choices were influenced by experience in concrete policy relevant activities. All participants were voluntarily involved in the workshops, were provided with information about the project prior to the workshop, and provided their written informed consent for all parts of the data collection process. All steps were taken to ensure participants' anonymity and confidentiality.

### 3.4 Data analysis

Data were transcribed verbatim and analysed using the qualitative analysis software NVivo Pro 12, using detailed coding schemes. The coding schemes identified the situations that may lead to misinformation and the actions that the stakeholders take as a result, and gathered information on how stakeholders viewed their own online practices relating to addressing misinformation, their perceived challenges and needs, and the type of evidence needed to trust the information. The analysis of the data adopted a grounded theory methodology (Strauss & Corbin, 1994) which, through the constant comparative method, seeks to develop theory using empirical evidence derived through a systematic analysis of the data. After initial discussions of the coding schemes and joint coding of parts of the data, inter-rater reliability was assessed by two of the authors using a new subset of the data. There was high consistency between the two raters, with Cohen's  $\kappa=0.863$  (95% CI, .737 to .988),  $p<.0005$ , which according to Landis and Koch (1977) is excellent. The two raters then independently coded the data set; these codes were also reviewed by two other senior members of the authors' team, thus increasing the reliability of the reported findings.

## 4. Findings

The qualitative analyses of the data corpus amassed from the focus group and in-depth individual interviews suggest continuity between the principles of traditional policy-making and data gathering online, but also emphasise the need for changing the internet platforms' landscape to adjust to the complexities of contemporary policy-making and the multivocality of engagement in democratic societies. Results suggest a need to introduce new platform policies and functionality, but also a need for renegotiating, re-prioritising, and re-inventing policy-making processes and key actors. Study participants called for action on two levels: immediate action level and long-term action. In terms of the immediate action level, there is a recognition that agents of policy change (such as trusted policy networks

and knowledge brokers) are still a vital aspect of the tradition of policy-making (Smith & Katikireddi, 2013), but also seem to be side-stepped by the rapid propagation of online information, as we will discuss next. Therefore, there is a need for reconceptualising practices that can support policy-making in the fast-paced, on-line world. In fact, results suggest that policy-making may already happen through technology mediation and algorithmic manipulations; study participants raise this as both beneficial and as a problem, and ask that human involvement assumes an equally significant role in the online policy-making process.

Stray's (2017) strategies to counter disinformation are aligned with policymakers' descriptions of actions they would like to take in response to misinforming posts on social media: they wish to refute, expose inauthentic comments, and provide alternative narratives. However, before they are able to do so, policymakers recognise that they, too, are uncertain of which information and whom or what they can trust, as verification measures traditionally in place are now collapsing or unable to support them in real-time: can they trust information they have in front of them? Who should help them interpret it? How can they transform their network to quickly fact-check and assess the reliability of the information in this new situation?

In the sections that follow, we first provide an overview of the challenges that were identified by the policymakers who participated in this study, and then present the four main themes which have implications for platform policies and contemporary view of policy-making in the context of combating misinformation: a) creating a trusted network of experts and collaborators, b) facilitating the validation of online information, c) providing access to visualisations of data at different levels of granularity, and d) increasing the transparency and explainability of flagged misinformative content. We conclude each section presenting the policymakers' viewpoints on each challenge with a brief commentary.

#### **4.1 Challenges for policymakers on social media**

The first goal of this study was to explore the policymakers' self-identified challenges in combating misinformation on social media. Policymakers identified several challenges, including lack of resources to validate information and correct misinformation, and their own professional routines. Among the chief challenges mentioned by policymakers was the amount of information that needs to be evaluated when formulating a decision. This makes validating the information both a time-consuming and a very complex process, and as a result, due to time constraints, there is often reliance on the expertise and knowledge of other peer deci-

sion-makers. The impact that misinformation can have on decision-making is a concern to policymakers, especially considering that the results of any decisions that may stem from misleading or inaccurate data may not be immediately visible. When faced with misinformation, time is of essence, and the need for immediate corrections or refutations is further complicated by the complexity of performing thorough information validation procedures.

Currently, the actions that policymakers can take online are limited, and post-specific reporting actions were either viewed as futile (i.e., having limited impact on the wider challenge of misinformation) or seen as potentially dangerous. Our study participants provided some examples on dilemmas they faced in their own line of work, where they feared that removals of what they perceived as extremist media posts could have a counter-effect and increase polarisation, as such censoring actions may be perceived as limiting the public's freedom of speech. Policymakers also expressed skepticism about the potential of technological solutions, stating that misinformation complexity requires critical reasoning by humans rather than automated solutions.

## **4.2 Reconceptualising social media platforms to create a trusted network of experts and collaborators**

Policymakers indicated that internet platforms do not currently support the collection of information that would be useful to a policymaker nor do they facilitate access to trusted networks and information online. Data from the focus group and the in-depth interviews clearly suggest that the study respondents already have existing networks of experts with whom they collaborate offline, and daily evaluate claims relating to news stories or posts. Streamlining these procedures and sharing knowledge on news stories that have been fact-checked internally (within a policymaker's organisation) or externally (by fact-checking organisations or journalists) appears to be a need that should be addressed. While still of use to policymakers, most of the policies already in place on social media target civil society; furthermore, there is still much that needs to be developed even for civil society to increase everyday users' resilience to misinformation. We explain the major themes regarding the identified needs and challenges relating to policymakers' use of the internet to monitor, control and reduce online misinformation.

Our analysis revealed a need for functionalities (such as dashboards connected to social media platforms) that would enable policymakers liaise with experts (i.e., subject-matter experts) and external collaborators (i.e., journalists or fact-checkers). Policymakers involved in the co-creation workshops had disparate fields of

professional expertise and expressed an interest in having the ability to define topics that are aligned with their own professional interests and expertise when addressing potential misinformation found online to allow for a speedier and more streamlined response.

The ability to quickly share internal decisions about misinformation correction with third parties such as legal counsels, was also raised as an important issue. Policymakers also commented on the importance of providing flagging functionalities, to enable them to tag suspected misinformation posts or news stories to be reviewed by fact-checkers, or to report posts to administrators claiming a post as misinformation based on their own professional expertise.

Integration with existing social media platforms would enhance the usability of any such tool. A tool that would enable and facilitate access to journalists was also considered important, with many policymakers emphasising the necessity of interdisciplinary synergies to harness collective expertise. Policymakers clearly expressed that there is limited access to journalists and pointed to the need to bridge this gap to allow for greater collaboration either by allowing tagging posts for review and assigning external evaluators, or by connecting them to other relevant parties (other policymakers, media officials, journalists, fact-checkers). Enabling policymakers to connect with fact-checkers, other experts, and other stakeholders to request or receive information relating to tagged news stories or posts, is an important design recommendation that emerged from our analyses. Further to this, facilitating this connection provides the opportunity of direct access to publishers and journalists and thus enables a speedier notification of publishers and journalists regarding corrected information. In addition, policymakers mentioned that there is a communication problem within institutional hierarchies, often due to limited access to others; therefore, a collaboration platform could also facilitate intra-organisational collaboration in terms of information sharing, or sharing of expertise, especially in larger, more complex institutions. As shown in the following excerpt, ideally, an online platform would facilitate communication across stakeholders and could also facilitate the fast dissemination of corrected information.

For governmental institutions, the means through which serious news that may have a negative result is addressed, is through a press release, which has to be released and disseminated fast and it is through this that all relevant publishers and institutions should be informed. The way that this press release is managed, doesn't have to be in the traditional way, like most press releases, but

it could take place through some kind of software, that can provide an immediate update. (Greece, Policymaker, in-depth interview, Co-creation Workshop 1)

Our analyses indicated that policymakers have an existing offline network of reliable and trustworthy experts or collaborators, and that providing policymakers with a tool that allows them to create a central database, which they can use to gain immediate feedback on the evaluation or correction status of misinforming content would be of particular benefit to them. This need to take additional actions online is illustrated in the following excerpt from an Austrian policymaker:

*“What is disseminated in the media, is useless. I would like to add to the authors. If necessary, inform and encourage people who know better to do something. Forward to multipliers. Or to experts.”* (Austria, Policymaker, in-depth interview, Co-Creation Workshop 1)

The potential for expediting the correction procedure, by centrally assigning and receiving feedback on disputed information was clearly presented as an appealing, if not required, response to the challenge of misinformation, especially provided that time is of the essence in issuing corrections online (Walter & Tukachinsky, 2020). The following excerpt illustrates the wishlist of a Greek policymaker considering how to correct misinformation regarding the monetary allowances that immigrants receive in Greece (a recurrent theme of heated discussions on social media).

I would like to have a button, not a ‘divine’ button that tells the truth, more like a procedure. I mean, a button sounds a little Orwellian, having a supreme force tell me if it’s all truth or all lies. But, have a fact-checker, to tell, if nothing else, if in the country there are 18 million illegal immigrants and if they get an allowance or not. These two pieces of information can be checked. (Greece, Policymaker, in-depth interview, Co-Creation Workshop 1)

### Commentary

Drawing from their experiences of facing the impact of misinformation, especially those relating to immigrants in Greece and Austria, the policymakers fall back to their everyday offline routines, which they would like to see extended online as they decide how to address misinformative comments. Their approach is, at least at the surface level, aligned with the criteria suggested by Vraga and Bode (2020),

and acknowledges that, on many occasions, they need to check in with experts, such as scientists or fact-checkers, in order to ascertain if something is indeed misinformation. Even though they do not discuss the dynamic and contextualised nature of misinformation, the need to triangulate their assessments, and reach evidence-based decisions is an important element of policy-making (Décieux, 2020). Future studies could examine how policymakers reason with evidence with the help of such trusted sources and provide more insight into technologically-mediated policy-making versus the respective conventional practices.

### **4.3 Facilitating the validation of online information in real-time**

Policymakers recognised that doctored information might be difficult to discern with human inspection alone and as a result expressed self-doubt about the evaluation of online information. Policymakers' proposed ideal solutions, and their responses in relation to the limitations to online actions they adopt when addressing misinformative content, led to a set of design recommendations that may inform platform policies for timely information validation. They suggested that online platforms can support this aspect of their work by providing tools to facilitate the policymakers' evaluation of online information. Some of these tools may help them automatically detect misinformation, while other functionality might help them report misinformative content or ask for expert help in reviewing it. Time pressure, and the fast news cycle pace, were flagged as important; delayed response to misinformation may have dire consequences for critical issues, such as violent actions and reactions to them. An example of the latter is, for instance, misinformation surfacing on Twitter and Facebook in late May/early June 2020, claiming that the police officer, charged with the death of George Floyd in the United States, was an actor and that the "incident had been faked by the deep state" (Alba, 2020); the spread of such misinformation can incite more violence against an already troubled state and its officers.

Policymakers could contribute to the validation of online information in multiple ways: for instance, they could use their own professional expertise to rate the trustworthiness of images or news story content, assess a news story's risk for spreading misinformation (for instance, high, medium or low risk), or add notes and context relating to identified misinformation. Designing for crowdsourcing the validation of news shared online, and allowing policymakers to categorise news organisations according to their level of trustworthiness based on their experiences, emerged as a design recommendation amongst participants. The following excerpt presents such a proposal by a policymaker in Greece:

I'm thinking now of some lists, like, for instance, white, grey and blacklists, which evaluate technological solutions' interoperability and electronic transaction services. It wouldn't be so bad if we had similar lists for some sites that provided information, that could inform the [potential] software or inform us. (Greece, Policymaker, in-depth interview, Co-Creation Workshop 1)

Displaying a social media post's reliability based on a set of specific criteria was a design recommendation that emerged across all three co-creation sites and in all focus groups. Differentiating amongst users' reliability online was also mentioned, with suggestions for differentiating posts shared by decision-makers, especially for posts that comment on misinformation and aim to raise awareness against misinforming content.

Providing the necessary support that reinforces the reliability and trustworthiness of sources was raised as a very important issue. It was also underlined that corrections provided by "verified" users should provide links to an explanation, or provide links to official responses, especially when a misinformation post is shared. Further to this, policymakers expressed more trust in established news organisations, and it was recommended that links either to related stories, or other news organisations' story on the same topic, as well as other official sources, should also be connected to potentially misinforming posts to enable information validation.

The process of validation would further be facilitated by providing information on the number of verified news organisations that have shared the same information. Participants noted that the identity of a news organisation reporting a story is important, but so is the journalist—therefore, it is necessary to provide sufficient information on the news institution's background, as well as on the journalist's professional experience and reputation to determine whether a piece of information can be trusted. This idea that the reporting of reputation should go beyond the official news source is evident in the following excerpt:

So, if he's an opinion leader or a journalist, who is himself spreading, in your own estimate, false news -- when I say false news, I mean in the sense of propaganda, false evidence, for instance wrong numerical figures -- then he is deconstructing himself. That is to say, if he's someone important, a journalist or a politician for instance, and he starts reproducing completely fake news, then at some point his own credibility falls, because it becomes something known. (Greece, Policymaker, in-depth interview, Co-Creation Workshop 1)

Even though policymakers indicated an awareness of existing tools and fact-checking resources, providing a repository of existing tools on social media platforms emerged as an important recommendation as none of the existing platforms or software plugins have all of the functionality they need. Participants also discussed the possibility of automating the rating of the veracity of the news automatically, as they mentioned that when assessing information online they also pay importance to the source itself, and specifically the source URL. They suggested that automation processes would be helpful in quickly sifting through information quality on social media, such as scanning the daily news reports to provide a summary, and potentially serve as a daily baseline. Since misinformation also includes pairing out-of-context images and news content or headlines, automating checks that evaluate image-news congruity was another design recommendation that stemmed from the policymakers' discussions. Finally, the proliferation of online bots and trolls was connected to this need and the ability to tag specific social media accounts as trolls, based on online behaviour. This would contribute to identifying and avoiding such accounts; removing or blocking bot or troll accounts on social media was also emphasised.

### Commentary

While policy-making includes slower paced decision-making, it also often includes the need for quicker decision-making to respond to smaller or larger crises. In a large crisis situation, in particular, it is important to respond quickly and accurately, hence the participants' recommendations for automating some of the verification processes to validate sources, filtering out unwanted information and reducing information overload. The policymakers did not substantially discuss the feasibility of such operations nor did they bring up the ethical aspects of the suggested automated processes. Given the definition of misinformation as a fluid, changing and dynamic body of knowledge, as described in Vraga and Bode (2020), and the risks that are inherent with censoring information, as a project we adopt the idea that computational tools can support quicker responses to addressing potentially misinformative content, but should not replace people's critical examination of the data. A current debate in the literature is whether *free speech* is increasingly being used as a shield to allow hate speech (Titley, 2020). Therefore, the design of such technical efforts should be done in a transparent and retraceable manner. Finally, we also believe that it is important for future work to address policymakers' mental models and operationalisation of misinformation, as well as provide case studies of how policymakers react to misinformation in real time.

## 4.4 Providing access to visualisations of data at different levels of granularity

Presently, internet platforms do not offer the functionality of different viewpoints into users' online activity. While big data are being amassed, they are mostly used for commercial, profit-making purposes. Policymakers in this study indicated that access to the information source trail, that is, mapping the information path and its spread on social media and providing ratings of user behaviours, as they relate to misinformation-sharing trends, would be useful. In addition, having the access to data that relates to the number of fact-checks or official refutations a post has amassed, would further enhance the post's credibility. Policymakers emphasised the importance of having accurate data in order to make decisions and mentioned that being able to view customised data at bird's eye view in regard to misinformation, i.e., within their country, would be particularly useful in assessing misinformation impact at a local level. For example, a Swedish policymaker asked for the following:

*"There should be a statistic on how often fake news is spreading over, at least, Scandinavia, or Sweden if you have it, but feels like it's not that much."* (Sweden, Policymaker, in-depth interview, Co-Creation Workshop 1)

Policymakers wished that social media platforms could provide access to official information, such as statistics, as it is easy for someone to misrepresent data. Even though data are available from official data sources, i.e. from national institutions, or at a European level from organisations such as Eurostat, including links to such official data sources, which are considered trustworthy by policymakers, would support evidence-based policy-making and enhance the trustworthiness of the information.

### Commentary

The policymakers in our study point to the usefulness of trusted sources of information and the power that easily accessible and dynamic representations could also provide to their decision-making. The issue of trustworthiness of the source of information or of the source of assessing the information has been a recurring theme in our conversations with policymakers and it is one that should be attended to when designing technological solutions for policymakers.

## 4.5 Increasing transparency and explainability of flagged misinformative content

Policymakers in our study frequently pointed out that citizens should also be part

of the effort to identify and curtail misinformation, and suggested a number of approaches and software features that could enhance misinformation resilience for everyday users. For instance, when misinformation is corrected, a post on social media that contains a link to a news story with false information should also include information on fact-checked data and provide access to officially corrected information.

The algorithmic regulation of online information can lead to technology operating as a non-human policymaker (Lessig, 1999; Just & Latzer, 2017), through filtering or blocking information from groups of users, who may often be unaware of such practices. Policymakers seem to frown upon such practices, as they perceive transparency as a facilitator of modern democracies (Hale, 2008). It is often reported that the public should have sufficient information to understand how algorithms manage media content (Picard & Pickard, 2017). At the moment, internet platforms do not provide sufficient information for citizens or policymakers to help them understand such manipulations. Some strategies seem to be more effective in correcting misperceptions, such as providing an explanation rather than a simple refute, exposing users to related but disconfirming stories, and revealing the demographic similarity of the opposing group (Nyhan & Reifler, 2015).

The importance of communicating the automation rules and maintaining transparency in how a technology that addresses misinformation may be automating certain procedures, remained a key concern for participants; this has also been discussed in the literature (see Butler, Joyce, & Pike, 2008). Providing access and facilitating communication procedures emerged as key desired components of any future technological solutions. Specifically, the policymakers indicated a need for tools to be open access and to provide clearly transparent procedures for their functionalities. This design recommendation was highlighted in the comment by a Swedish policymaker, as shown below:

*“A lot of these methods, they aren’t open access so you cannot see how they are operating and if these methods of fact-checking are actually accurate. So, it’s also very hard to tell you can’t judge the whole truth from these kinds of sites either because of that.”* (Sweden, Policymaker, in-depth interview, Co-Creation Workshop 1)

At the same time, several policymakers across sites suggested a need for a more media-literate citizenry, which can question and engage in critical reflection about the information it is receiving on social media.

## Commentary

Transparency and explainability are rising in importance as artificial intelligence, algorithms and recommender systems are increasingly being used to personalise the online experience and to promote commercial and other interests. According to Llansó, van Hoboken, Leerssen, and Harambam (2020) policymakers' possible interventions to recommender systems' actions include "content removal and other forms of moderation, algorithmic content curation, user customization options, transparency, and media literacy" (p. 18). The policymakers' suggestions were aligned with these recommendations and showed awareness of the issues. At the same time, the policymakers' responses indicated an awareness of the problematic aspects when it comes to implementing these technologies. They appear to understand that the use of powerful technologies also requires technology platforms to be accountable by becoming more transparent in how they work, and a media literate populace. These discussions indicate the complexity of the issue and reiterate the importance of synergies across time, scale and societal actors. The data we collected have, nonetheless, remained at a more general level; it is important that future work delves deeper into what constitutes acceptable and satisfactory levels of explainability and transparency on social media and for whom.

## 5. Conclusions

This paper discussed how policymakers view their role in combating the spread of misinformation on social media, and their self-identified challenges and proposed possible solutions to address them. Key findings included the need for access to trusted networks online, for technological resources to quickly identify and address misinformation, making different levels of data representations available to support evidence-based decisions, and for increased transparency and explainability of misinformation-targeting actions online.

Policy networks and knowledge brokers have always been important in policy-making (Agranoff & McGuire, 2001). Gathering and activating resources, including social and human capital, is key to successful policy networks. Although there is considerable work in successfully managing a network that can guide effective policy-making, our study participants indicate that there is a need for re-imagining social media platforms' facilitation of policy-making to address online misinformation. Renegotiating policy-making spaces refers to providing policymakers with new tools and providing them with new ways to access policy networks online (as compared to more traditional, slower methods for such access). These new spaces and modes of collaboration should be dynamic, re-configurable and accessible on-

demand.

The practical implications of policymakers' self-identified needs may be interpreted as requests for social media platforms that cater to specific decision-making needs. Requests such as facilitating collaboration with experts and journalists, providing the opportunity to solicit fact-checks and report problem cases (such as the possibility of trolls, blatant misinformation, etc.), as well as allowing for personalisation and customisation across different socio-cultural contexts, were reported by policymakers as deviating from social media's current scope. Such changes would require dashboards that allow intra-organisational collaboration, as well as enable inter-organisational synergies among journalists, fact-checkers, and other external collaborators within policymakers' trusted networks. Are internet platforms interested and capable of instigating these kinds of changes? Do they see governmentality (Foucault, 1991), that is societal action that is de-centralised and can also be enacted by citizens, as part of their mission and values? Or are these developments part of the struggle between platforms and others as in terms of power and control?

Platform policies can be fully automated, such as evaluating the credibility of the content by using misinformation detection algorithms, or can be semi-automated when fully automated policies cannot be applied and human intervention is required for a policy to be supported by automation. Examples of the latter include reporting misinforming content that was not clearly identified by automatic algorithms or reporting content that was mistakenly identified as misinformation. Such policies are semi-automated since they need manual intervention from users to provide evidence, such as a URL or other feedback for the post that may have been inaccurately flagged or labeled. In ongoing work, we are developing and investigating a software plugin that implements such policies. We are also developing a social media dashboard prototype for policymakers to examine how such software can address some of the challenges on evidence-based approaches to decision-making and combating misinformation on social media. In addition to socio-technical solutions that can address some of the gaps of knowledge in contemporary policy-making, more research is needed that can illustrate the impact of such tools, and present case studies and data on the potential of the contribution of citizen participation in policy-making.

We ultimately argue that this reconceptualisation of social media platforms is a necessary adaptation in the current complex online information environment; the speed and quantity of information on social media make it difficult to monitor, flag and curb misinformative content, and policymakers have little by way of combat-

ing misinformation online. Indeed, this can sometimes even lead to the inadvertent spread of misinformation by key policy actors themselves, especially in information-rich and diverse situations (Brennen et al., 2020). While recommendations for revised platform policies may still appear as a wish list, the prominence of social media in the information and news ecosystem necessitates that these are taken under serious consideration.

---

## ACKNOWLEDGEMENTS

This work has been partially funded by the Co-Inform project (770302), under the Horizon 2020 call “H2020-SC6-CO-CREATION-2016-2017 (CO-CREATION FOR GROWTH AND INCLUSION)” of the European Commission. The authors acknowledge the support of the Co-Inform team and partners for the organisation of the co-creation workshops and the data collection efforts in Austria, Greece, and Sweden. Special thanks go especially to the following colleagues who have contributed to the data collection and analysis efforts: Vasilis Peristeras and Nancy Routzouni, at *International Hellenic University*; Somya Joshi, Mohamed Saqr, Oxana Casu, and Dimitris Sotirchos, at *Stockholm University*; Ipek Baris and Oul Han, *Universität Koblenz-Landau*; Steffen Staab, *University of Stuttgart*; and Love Ekenberg, *IASA*. More information on Co-Inform can be found at: <https://coinform.eu/>

---

## References

- Agranoff, R., & McGuire, M. (2001). Big questions in public network management research. *Journal of Public Administration Research and Theory*, 11(3), 295–326. <https://doi.org/10.1093/oxfordjournals.jpart.a003504>
- Alba, D. (2020, June 1). Misinformation about George Floyd protests surges on social media. *The New York times*. <https://www.nytimes.com/2020/06/01/technology/george-floyd-misinformation-online.html>
- Avery, E. J. (2017). Public information officers’ social media monitoring during the Zika virus crisis, a global health threat surrounded by public uncertainty. *Public Relations Review*, 43(3), 468–476. <https://doi.org/10.1016/j.pubrev.2017.02.018>
- Braman, S. (2016). Instability and internet design. *Internet Policy Review*, 5(3). <https://doi.org/10.14763/2016.3.429>
- Brennen, J. S., Simon, F., Howard, P. N., & Nielsen, R. K. (2020). *Types, sources, and claims of COVID-19 misinformation* [Factsheet]. Reuters Institute for the Study of Journalism. <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>
- Butler, B., Joyce, E., & Pike, J. (2008). Don’t look now, but we’ve created a bureaucracy: The nature

and roles of policies and rules in Wikipedia. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1101–1110. <https://doi.org/10.1145/1357054.1357227>

Cuan-Baltazar, J. Y., Muñoz-Perez, M. J., Robledo-Vega, C., Pérez-Zepeda, M. F., & Soto-Vega, E. (2020). Misinformation of COVID-19 on the Internet: Infodemiology study. *JMIR Public Health and Surveillance*, 6(2), 18444. <https://doi.org/10.2196/18444>

Décieux, J. P. P. (2020). How much evidence is in evidence-based policymaking: A case study of an expert group of the European Commission. *Evidence & Policy: A Journal of Research, Debate and Practice*, 16(1), 45–63. <https://doi.org/10.1332/174426418X15337551315717>

European Commission. (2018). *Media Use in the European Union. Autumn 2017* (Report No. 88; Standard Eurobarometer). Publications Office of the European Union. <https://doi.org/10.2775/116707>

Fischer, F. (2003). *Reframing public policy: Discursive politics and deliberative practices*. Oxford University Press. <https://doi.org/10.1093/019924264X.001.0001>

Flynn, D. J., Nyhan, B., & Reifler, J. (2017). The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics. *Political Psychology*, 38(51), 127–150. <https://doi.org/10.1111/pops.12394>

Foucault, M. (1991). Governmentality. In G. Burchell, C. Gordon, & P. Miller (Eds.), *The Foucault effect: Studies in governmentality* (pp. 87–104). University of Chicago Press.

Freeman, R. E., Harrison, J. S., Wicks, A. C., Parmar, B. L., & De Colle, S. (2010). *Stakeholder theory: The state of the art*. Cambridge University Press.

Fung, A. (2013). The Principle of Affected Interests: An Interpretation and Defense. In J. H. Nagel & R. M. Smith (Eds.), *Representation: Elections and Beyond*. University of Pennsylvania Press. <https://www.jstor.org/stable/j.ctt3fhtrg>

Graves, L. (2018). *Understanding the Promise and Limits of Automated Fact-Checking* [Factsheet]. Reuters Institute for the Study of Journalism. <http://www.digitalnewsreport.org/publications/2018/factsheet-understanding-promise-limits-automated-fact-checking/>

Hale, T. N. (2008). Transparency, accountability, and global governance. *Global Governance: A Review of Multilateralism and International Organizations*, 14(1), 73–94. <https://doi.org/10.1163/19426720-01401006>

High Level Group on fake news and online disinformation. (2018). *A multi-dimensional approach to disinformation: Report of the independent High level Group on fake news and online disinformation* [Report]. Publications Office of the European Union. <https://doi.org/10.2759/739290>

Innes, J. E., & Booher, D. E. (2003). Collaborative policymaking: Governance through dialogue. In M. Hajer, M. A. Hajer, H. Wagenaar, R. E. Goodin, & B. Barry (Eds.), *Deliberative policy analysis: Understanding governance in the network society* (pp. 33–59). Cambridge University Press.

Just, N., & Latzer, M. (2017). Governance by algorithms: Reality construction by algorithmic selection on the Internet. *Media, Culture & Society*, 39(2), 238–258. <https://doi.org/10.1177/0163443716643157>

Kapp, J. M., Hensel, B., & Schnoring, K. T. (2015). Is Twitter a forum for disseminating research to health policy makers? *Annals of Epidemiology*, 25(12), 883–887. <https://doi.org/10.1016/j.annepidem.2015.09.002>

- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Larson, H. J. (2020). Blocking information on COVID-19 can fuel the spread of misinformation. *Nature*, 580(7803), 306. <https://doi.org/10.1038/d41586-020-00920-w>
- Lessig, L. (1999). *Code: And other laws of cyberspace*. Basic Books.
- Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131. <https://doi.org/10.1177/1529100612451018>
- Llansó, E., Hoboken, J., Leerssen, P., & Harambam, J. (2020). *Artificial Intelligence, Content Moderation, and Freedom of Expression* (Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression Series) [Working Paper]. Annenberg Public Policy Center. [https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/06/Artificial\\_Intelligence\\_TWG\\_Llanso\\_Feb\\_2020.pdf](https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/06/Artificial_Intelligence_TWG_Llanso_Feb_2020.pdf)
- Magalhães, J. C., & Katzenbach, C. (2020). Coronavirus and the frailness of platform governance [Op-Ed]. *Internet Policy Review*. <https://policyreview.info/articles/news/coronavirus-and-frailness-platform-governance/1458>
- Nyhan, B., & Reifler, J. (2015). Does correcting myths about the flu vaccine work? An experimental evaluation of the effects of corrective information. *Vaccine*, 33(3), 459–464. <https://doi.org/10.1016/j.vaccine.2014.11.017>
- Perrin, A., & Kumar, M. (2019, July 25). About three-in-ten U.S. adults say they are 'almost constantly' online [Blog post]. *Pew Research Center, Fact Tank*. <https://www.pewresearch.org/fact-tank/2019/07/25/americans-going-online-almost-constantly/>
- Picard, R., & Pickard, V. (2017). *Essential principles for contemporary media and communications policymaking*[Report]. Reuters Institute for the Study of Journalism. <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/research/files/Essential%2520Principles%2520for%2520Contemporary%2520Media%2520and%2520Communications%2520Policymaking.pdf>
- Renn, O. (2008). *Risk governance: Coping with uncertainty in a complex world*. Routledge. <https://doi.org/10.4324/9781849772440>
- Smith, K. E., & Katikireddi, S. V. (2013). A glossary of theories for understanding policymaking. *Journal of Epidemiology & Community Health*, 67(2), 198–202. <https://doi.org/10.1136/jech-2012-200990>
- Strauss, A., & Corbin, J. (1994). Grounded theory methodology. In K. D. Norman & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273–285). Sage Publications.
- Stray, J. (2019). Institutional Counter-disinformation Strategies in a Networked Democracy. *Companion Proceedings of the 2019 World Wide Web Conference*, 1020–1025. <https://doi.org/10.1145/3308560.3316740>
- Titely, G. (2020). *Is Free Speech Racist?* John Wiley & Sons.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Vraga, E. K., & Bode, L. (2020). Defining misinformation and understanding its bounded nature: Using expertise and evidence for describing misinformation. *Political Communication*, 37(1),

136–144. <https://doi.org/10.1080/10584609.2020.1716500>

Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it? *Communication Research*, 47(2), 155–177. <https://doi.org/10.1177/0093650219854600>

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making* (Report DGI(2017)09). Council of Europe. <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c>

Published by



ALEXANDER VON HUMBOLDT  
INSTITUTE FOR INTERNET  
AND SOCIETY

in cooperation with



CREATE

centre  
— internet  
— et  
— societe



R&I

IN3  
Internet  
interdisciplinary  
Institute

Universitat Oberta de Catalunya