Disinformation during the digital era: A European code of self-discipline

Paolo Cesarini,
European Commission

Abstract:
Given the strategic role acquired by the social media as the preferred channel of information and the changes wrought by digital technology in the news media, fake news can circulate more easily. It now has unprecedented circulation and penetration rates, and societal impact. Efforts at the European level are described for stymying this phenomenon, in particular, the self-discipline foreseen by the "EU Code of practice on disinformation" proposed by the big online platforms in October 2018. How effective are these measures in relation to the five types of vulnerability typical of the media? How can they cope with the methods whereby hostile forces, whether governmental or not, manipulate information for the purpose of making a profit or of political subversion?

The phrase "fake news" has intruded upon everyday language since 2016 — the American presidential campaign and Brexit referendum.[1] More recent operations of disinformation in Europe and elsewhere have added to its disturbing popularity.[2] Another example is the hybrid attack targeting the last phase in the 2017 presidential campaign in France. The March 2018 attack against the former Russian intelligence officer, Sergei Skripal, in Salisbury, UK, has illustrated the diplomatic and geopolitical consequences of an orchestrated diffusion of fake news, in particular on the social networks.

To deal with these threats, the European Council and Parliament, as well as organizations representing civil society and the media, urgently called for a suitable response. According to a recent Eurobarometer opinion poll, 83% of Europeans see fake news as a menace to democracy that is widespread in all EU member states.[3] As the European Commission pointed out in a communication, disinformation undermines citizens' trust of public institutions and the media, both online and traditional, because it attacks democratic values (EC 2018b). By altering the capacity of persons to form an opinion and make decisions with full knowledge of the facts, it restricts free speech. When circulated on a large scale, disinformation can skew public opinion about topics as important as immigration, climate change or health, jeopardize internal security or subvert the integrity of elections.

The operators of online services (in particular, Facebook, Google and Twitter) that allow fake news to be rapidly spread and precisely targeted as never before to so many cybernauts must now face up to their social responsibility. The report to the EC in March 2018 by a high-level group of

[1] The views expressed in this article are the author's own. This article has been translated from French by Noal Mellott (Omaha Beach, France). The translation into English has, with the editor's approval, completed a few bibliographical references. All websites have been consulted in July 2019.

[2] FREEDOM HOUSE "Freedom on the Net 2017: Manipulating Social Media to Undermine Democracy" available at https://freedomhouse.org/report/freedom-net/freedom-net-2017.

[3] Eurobarometer (2018) "Final results of the Eurobarometer on fake news and online disinformation", 12 March at https://ec.europa.eu/digital-single-market/en/news/first-findings-eurobarometer-fake-news-and-online-disinformation. The full report is available at http://ec.europa.eu/commfrontoffice/publicopinion/index.cfm/ResultDoc/download/DocumentKy/82797.

experts has underscored the complexity of this phenomenon and the need for a multidimensional approach with actions on several fronts that implicate all parties concerned, both public and private (EC 2018a). The aforementioned communication, which closely followed up on the recommendations made by this group of experts, defined several complementary and interdependent actions for holding platforms responsible, backing an independent network of fact-checkers Europewide, promoting media literacy, supporting a journalism of quality, and, at the same time, enhancing the resilience of electoral processes to cope with cybermenaces.

The approach retained was based on the principle of self-regulation. A concrete outcome of applying this principle is the "EU code of practice on disinformation" of September 2018.[4] Established by a multiparty forum involving technological firms, the advertising industry, media and civil society organizations, this code is the first example worldwide of self-regulation in this field. It is open to all stakeholders. In October 2018, Twitter, Facebook, Google, Mozilla and several European associations representing the advertising industry signed it.

To gauge the scope and potential impact of the "EU code of practice on disinformation", let us start by looking at the typical vulnerabilities in the current system.

Systemic vulnerabilities

Several recent studies have suggested that the digital transformation of the media and the rise of online platforms are at the origin of five major systemic flaws that hostile parties can put to use in pursuit of for-profit goals or as a tool of political subversion (CHRISTIE 2018, KAVANAGH & RICH 2018, JEANGÈNE-VILMER *et al*. 2018, MATZ *et al*. 2017, WARDLE & DERAKHSHAN 2017, DEL VICARIO *et al*. 2016):

- **MICROTARGETING AND PERSONALIZATION OF ADVERTISEMENTS,** in particular political or activist ads. The increasing production, collection and analysis of huge quantities of personal data make a detailed psychometric profiling of users possible. Combined with the application of advanced techniques in predictive analytics and artificial intelligence, this abundance of data can serve to customize political advertising in order to finely target its distribution and increase its impact on vast audiences. Since users are not necessarily aware that their personal data have been used, they can be deceived about the nature or meaning of the information they receive. The Cambridge Analytica/Facebook scandal that broke out in March 2018 clearly demonstrates such a lack of transparency.
- **ASTROTURFING** refers to a set of malicious techniques, manual or algorithmic, for simulating the activity of a crowd on a social network. Operations by automatic systems (bots) and regiments of mercenaries (trolls) are coordinated to create a crowd effect on social networks by diffusing convincing spams and opening fake accounts in order to lend credit to (or make normal) contents that are extreme or even outright false. The goal is to make people believe that these polarizing messages come from varied sources and have wide support. Owing to their cognitive biases, users are induced to believe in these messages and share them on the social networks. This bolsters populist ideas and circulates fake news about sensitive topics, such as the effects of vaccination.
- **CLICKBAITING** is a vulnerability stemming from the current practices of online advertising. Ad agencies guarantee to place in real time advertisements that, thanks to automated decision-making processes, provide a pecuniary reward to the hosting website as a function of the number of times that visitors click on the ad. Headlines with sensational contents, including disinformation, are thus placed on a website with the intent to capture the attention of visitors

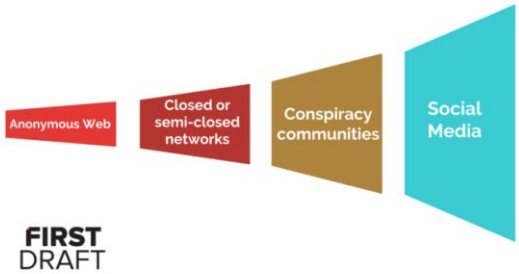
[4] See the EC's news article of 26 September 2018 at https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation. The "EU code of practice on disinformation" is available at https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54454.

and play on their emotions. A group of adolescents in Veles, Macedonia, actively used this technique for the purpose of turning a profit during the 2016 presidential American campaign.

● **ALGORITHMIC DISTORTIONS.** Digital platforms aggregate and filter contents by using algorithms that process cybernauts' digital tracks so as to make a granular analysis of their individual preferences. For a platform, the goal is to provide relevant, attractive contents in order to draw users' attention and keep them longer on the website. This increases the site's advertising potential (and eventually its income) by increasing the number of page views. The operation of these algorithms is murky however and, in some cases, skewed through a sort of "self-radicalization" that orients users to view increasingly extreme, tendentious or false contents (RIEDER *et al*. 2018).

● **DIFFUSION PROCESSES**. Finding out who is responsible for a disinformation campaign is complicated. Actions are seldom carried out by isolated individuals, but more often result from a coordinated effort by several parties acting at different levels in the digital ecosystem. Claire Wardle has recently uploaded a graphic of the chain of amplification of disinformation as it passes from an anonymous website through more or less closed networks and conspiracy communities onto the social media and is then taken up by the professional media (See Figure 1). As empirical research has shown "*Falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects were more pronounced for false political news*" (VOSOUGHI *et al*. 2018). People are more inclined to share fake contents on the social media. Users and even professional journalists and publishers (newspapers, broadcasting companies, etc.) might involuntarily or unconsciously amplify a fake information.

Figure 1: Diffusion of fake news.
Source: Claire Wardle on https://firstdraftnews.org/5-lessons-for-reporting-in-an-age-of-disinformation/.



The responsibility of the social media platforms

These vulnerabilities are decisive in the dissemination of fake news. One of the principal, oft mentioned causes of this problem is the absence of regulations about the social media's responsibility.

The EU directive on electronic commerce establishes a system of limited liability for digital platforms that provide services restricted to the hosting of third-party contents.[5] Such platforms are responsible for deleting (only) illegal contents fast, as soon as they become aware of them, and for adopting useful measures so that the contents not be reposted on the platform. The audiovisual services directive[6] approved on 14 November 2018 has modified this legislation by requiring that the social media adopt in advance good practices for both deleting illegal contents (such as posts promoting hatred, violence or terrorism) and restricting the access to deleterious contents (in

[5] For the directive's text: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32000L0031.

[6] For the directive's text: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018L1808.

particular for children). This directive has made legally binding the objectives to be reached on the European scale for optimizing the efforts already undertaken or to be undertaken by the platforms and member states in order to effectively address problems of this sort, which reach beyond national borders.

Fake news, which can be harmful without being illegal as such, is not covered by any specific EU regulation. This phenomenon's mutlifaceted nature makes it very difficult to legislate. Fake news can be an instrument used in a domestic political battle or a preferred means for orchestrating campaigns of hate in order to intimidate or slander public figures or social groups. It comes from forces inside or outside a country, forces that might be state-sponsored or not. However fake news might also crop up in a context of satire, parody or what obviously amounts to a form of social criticism. It might even result from journalistic errors. Given this diversity, a pure and simple requirement to delete fake news obviously risks interfering with the freedom of expression. In particular, Article 11 of the Charter of Fundamental Rights of the European Union provides for the rights of citizens "*to receive and impart information and ideas without interference by public authority and regardless of frontiers*".[7] It thus protects contents that are satires, parodies and legitimate social criticism independently of whether they are extreme or shocking.

In December 2018, the EC's "Action plan against disinformation" formulated a definition of disinformation with two essential aspects: "*Disinformation is understood as verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm*" (EC 2018c:1). Disinformation is qualified by the nature of its contents and the intention to deceive. A regulatory approach based simply on defining fake news as being harmful carries the risk of an Orwellian nightmare of a judge of truth, whether public or private. The element of intentionality eliminates this risk by characterizing disinformation in terms of behaviors harmful for the ecosystem of information and in terms of the vectors, technological tools and methods used. This category covers, for example, campaigns of disinformation intended to undermine the integrity of elections via abusive methods of communication. President Macron's bill of law, which criminalizes fake news under specific conditions (massive and artificial dissemination) and in the specific context of elections, seems to adhere to this rationale.

A code of self-discipline

In this context, the social networks and other parties responsible for the effects stemming from the aforementioned vulnerabilities must recognize the need to re-examine their practices so as to keep them from contributing to the amplification of disinformation and to better manage their technology in relation to the flaws in the communications system. For example, deepfakes, fake accounts or automated systems (using bots) can amplify the impact of fake contents on the public. Furthermore, such actions are behaviors with the evident intent to deceive. The platforms should, therefore, undertake corrective actions; and sanctions should be effective and proportionate if actions are not forthcoming.

The "EU code of practice on disinformation"[4] of September 2018 lists the five types of vulnerability that mostly account for online disinformation going viral. It commits the platforms to take relevant measures Europewide for correcting them, among them:

- measures of transparency about "*political and issue-based advertising*" so that sponsors are identified along with the amount paid and the criteria used for targeting. This entails setting up permanent digital archives to be made available for research.

[7] For the full text: https://en.wikisource.org/wiki/Charter_of_Fundamental_Rights_of_the_European_Union.

- measures for improving the resilience of services when actions (fake accounts, attacks from automated systems, astroturfing) try to radicalize and polarize communications.
- the development of tools for online advertisement placement that protect trade names and reduce the income of websites primarily financed through their promotion of fake news.
- algorithmic filters that reflect the reliability of sources more than the popularity of an information, and that automatically bundle and distribute a plurality of contents reflecting various viewpoints on wedge issues.

Finally, in order to better cope with the dissemination of disinformation, the EU code of practice asks platforms to make a commitment to work with research and fact-checking organizations, and accept that their contents be independently and constantly monitored on the European scale. This implies that they make accessible the data necessary for understanding risks and tracking viral contents during disinformation campaigns while upholding, of course, the General Data Protection Regulation's provisions on protecting personal data (GDPR).[8]

This code's effectiveness will mainly depend on whether the signatories strictly and persistently pursue its objectives. The efforts made will have to be on par with the size of each signatory and its responsibility. The implementation of these good practices will have to be regularly assessed by the EC in cooperation with member state's regulatory authorities by using transparent performance indicators. Tight monitoring is now under way.

Though necessary, the "EU code of practice on disinformation" alone does not suffice to neutralize disinformation in the current environment. For one thing, the effectiveness of the recommended good practices (*e.g.*, better understanding and controlling the momentum of dissemination or better protecting users) will depend on actions undertaken in parallel by the EC and member states — actions such as: supporting academic research and the emergence of an independent European network of fact-checkers, fostering media literacy among the public and supporting professional journalism. For another, factors unrelated to technology (such as direct interference by third-party states or the political polarization stemming from economic inequality or from forces that push toward radicalization and extremism in society) play a role just as critical. As a consequence, the EC's "Action plan against disinformation" of December 2018 has foreseen other interventions, in particular the creation of means of coordination between national authorities and EU institutions, and the creation of a rapid warning system so that the authorities in charge can more easily exchange information and better analyze threats (EC 2018c). The "EU code of practice on disinformation" is a pillar of these actions.

[8]The GDPR: "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data" available via: http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1478961410763&uri=CELEX:32016R0679.

References

CHRISTIE E.H. (2018) “Political subversion in the age of social media”, *Policy Brief*, October available via https://martenscentre.eu/sites/default/files/publication-files/ces_policybrief_political-subversion-v4.pdf.

DEL VICARIO M., VIVALDO G., BESSI A., ZOLLO F., SCALA A, CALDARELLI G. & QUATTROCIOCCHI W. (2016) “Echo chambers: Emotional contagion and group polarization on Facebook”, *Nature: Scientific Reports*, 6, article 37825, 14p. available at https://arxiv.org/pdf/1607.01032v1.pdf.

EUROPEAN COMMISSION (2018a) “A multi-dimensional approach to disinformation: Report of the High Level Expert Group on fake news and online disinformation” to the European Commission, 12 March, 44p. available via http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=50271.

EUROPEAN COMMISSION (2018b) “Tackling online disinformation: a European approach”, COM 2018/236/final, 26 April, available via https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0236.

EUROPEAN COMMISSION (2018c) “Action plan against disinformation”, joint communication of 12 May, 30p. available via https://ec.europa.eu/commission/sites/beta-political/files/eu-communication-disinformation-euco-05122018_en.pdf.

JEANGÈNE-VILMER J.B., ESCORCIA A., GUILLAUME M. & HERRERA J. (2018) *Les Manipulations de l’information. Un défi pour nos démocraties* (Paris: Centre d’Analyse, de Prévision et de Stratégie et Institut de Recherche Stratégique de l’École Militaire) available via https://www.diplomatie.gouv.fr/IMG/pdf/les_manipulations_de_l_information_2__cle04b2b6.pdf.

KAVANAGH J. & RICH M. (2018) *Truth Decay: An Initial Exploration of the Diminishing Role of Facts and Analysis in American Public Life* (Santa Monica, CA: RAND Corporation).

MATZ S.C., KOSINSKI M., NAVE G. & STILLWELL D.J. (2017) “Psychological targeting as an effective approach to digital mass persuasion”, *Proceedings of the National Academy of Sciences of the United States of America*, 114(48), pp. 12714-12719 available via https://www.pnas.org/content/pnas/114/48/12714.full.pdf.

RIEDER B, MATAMOROS-FERNANDEZ A. & COROMINA O. (2018) “From ranking algorithms to ‘ranking cultures’: Investigating the modulation of visibility in YouTube search results”, *International Journal of Research into New Media Technologies*, 24(1), pp. 50-68.

VOSOUGHI S., ROY D. & ARAL S. (2018) “The spread of true and false news online”, *Science*, 359, pp. 1146-1151 available at https://science.sciencemag.org/content/sci/359/6380/1146.full.pdf.

WARDLE C. & DERAKHSHAN H. (2017) *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making* (Strasbourg, FR: Council of Europe report DGI 09), 109p. available via https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c.