IN-DEPTH ANALYSIS

Requested by the IMCO committee

European Parliament

The legal framework to address "fake news": possible policy actions at the EU level

Policy Department for Economic, Scientific and Quality of Life Policies
Author: Andrea Renda (CEPS - Centre for European Policy Studies and College of Europe)
Directorate-General for Internal Policies
PE 619.013- June 2018

EN

The legal framework to address "fake news": possible policy actions at the EU level

Abstract

This paper argues that the current policy initiatives adopted by the European Commission are meaningful, but still incomplete. The policy response to online disinformation should ideally rely on: (i) the promotion of responsible behaviour in conveying information to end users; (ii) the enactment of a proactive media policy aimed at promoting pluralism and improving the exposure of diverse content to end users; and (iii) the empowerment of end users through media literacy initiatives, and supports to user behaviour.

This document was prepared by Policy Department A at the request of the Committee on the Internal Market and Consumer Protection.

This document was requested by the European Parliament's Committee on the Internal Market and Consumer Protection.

AUTHORS
Andrea Renda (CEPS - Centre for European Policy Studies and College of Europe)

ADMINISTRATOR RESPONSIBLE
Mariusz MACIEJEWSKI

EDITORIAL ASSISTANT
Irene VERNACOTOLA

LINGUISTIC VERSIONS
Original: EN

ABOUT THE EDITOR
Policy departments provide in-house and external expertise to support EP committees and other parliamentary bodies in shaping legislation and exercising democratic scrutiny over EU internal policies.

To contact the Policy Department or to subscribe for updates, please write to:
Policy Department for Economic, Scientific and Quality of Life Policies
European Parliament
B-1047 Brussels
Email: Poldep-Economy-Science@ep.europa.eu

Manuscript completed in June 2018

This document is available on the internet at:
http://www.europarl.europa.eu/supporting-analyses

DISCLAIMER AND COPYRIGHT

CONTENTS

LIST OF FIGURES

EXECUTIVE SUMMARY

Background

Due to the diffusion of digital technologies, access to news has become easier and cheaper than ever in history and news consumption has grown unprecedentedly. These positive developments were however accompanied by new problems, which have caught the attention of policymakers also due to recent scandals such as the alleged Russian meddling in national elections between 2016 and 2018, as well as Cambridge Analytica. Trust in online intermediaries suddenly plummeted, and Mark Zuckerberg's appearances before the U.S. Congress and the European Parliament marked the beginning of a new, possibly darker era in the relationship between governments and tech giants. This is the age of responsibility for companies that are as big as also relatively young, and admittedly were unprepared to handle all the consequences of their breath-taking growth.

As a matter of fact, online intermediaries and platforms have risen also thanks to a regulatory framework that shielded them from responsibility for editorial control and for possible misconduct by their users. This also translated into a massive flow of online news, which were subject to limited fact- and originality-checking. Importantly, in the Internet environment the correlation between news outreach and quality-checks is very limited: while in the pre-Internet age, only media outlets with sufficient resources could reach out to large audiences, today anyone can be a publisher with global reach. This also means that there may be limited incentives to invest in quality journalism, unless end users have a way to single out high quality from low-quality news. Moreover, the link between the content of the news and the means of communication is broken: as long as fake news bring traffic, money (i.e. from advertising), and no liability, there is no strong reason to expect that digital platforms will have a strong incentive to take independent action against disinformation.

The problem of quality of information in an information-rich environment such as the Internet led to the emergence of a variety of effects. These include:

- Content bubbles occur whenever an individual interacts with a single news source, powered by an algorithm that only feeds users based on their perception of what they will like, or be interested in. They are the product of both behavioural biases (such as the "confirmation bias", i.e. users tend to like what they already agree with); and the use of algorithms for personalized search, which are based on our past searches and thus mostly select content from a narrow subset of available sources.
- Unintentional fakes/opinions (so-called "misinformation"). Expressing one opinion, however false, can fuel disinformation even if there is no underlying intention to engage with manipulation of the public opinion. The Internet amplifies these statements, leading to an unprecedented rate of circulation of both true and false statements.
- Intentional fakes or amplifiers ("Disinformation"). In a subset of cases, the lack of filtering on online platforms has led to the abuse of such platforms, with the clear intention to manipulate public opinion, for example in the occasion of an election. Many of these news are ignored by the public as clearly fake, but others spread very quickly and can affect public opinion creating a thick layer of "noise" which causes trouble and confusion for end users.
- Disinformation (or influence) operations. These are commercially or politically motivated manipulation strategies, a variant of intentional disinformation, but happening at a much greater scale.

These operations, which can be state-sponsored, aim at affecting the outcome of elections or at discrediting commercial rivals by purchasing privileged spots for online advertisements and using them to spread intentionally and strategically crafted messages.

Not all these phenomena should lead to censorship or attempt to block the spread of information. Asking internet intermediaries to filter out non-majoritarian, non-fact-based opinions from the major channels of access to information would dramatically impoverish our democracy and society. The possible balance between freedom of expression and the right to be properly informed lie in: (i) the promotion of responsible behaviour in conveying information to end users; (ii) the enactment of a proactive media policy aimed at promoting pluralism and improving the exposure of diverse content to end users; (iii) the empowerment of end users through media literacy initiatives, and supports to user behaviour.

The European Commission has adopted initiatives in all of the three above-mentioned areas: they include setting up the High Level Group on Fake News and Online Disinformation; a Public Consultation on fake news and online disinformation; and the recent Communication on "Tackling online disinformation". The latter announces the creation of a multi-stakeholder forum, a study to examine the applicability of EU rules and possible gaps, the creation of an independent European network of fact-checkers, the launch of a secure European online platform on disinformation, and the promotion of voluntary online systems allowing the identification of suppliers of information based on trustworthy electronic identification and authentication means, including verified pseudonyms, as provided under the Regulation on electronic identification. This paper reaches the conclusion that the proposed measures would not be sufficient to promote a more sustainable evolution of the online news market: the policy mix would need to include a proactive media policy oriented towards "exposure" and pluralism; an ambitious plan to promote media literacy and user empowerment; and a forward-looking approach to innovative solutions for the era of "deep fakes".

1. INTRODUCTION

The course of history is often unpredictable, and things can turn around in unexpected ways. The past decades are full of examples: in the mid-1970s, as the oil crisis was raging, famous German economist Rudiger Dornbusch observed that "the crisis takes a much longer time coming than you think, and then it happens much faster than you would have thought"[1]. That is to say, very often unstable equilibria are sustained by the economy for a very long period, but then things collapse all at once, with no easy ways to mitigate the damages. The same, apparently, has happened to the Internet, or at least to end users' trust in the Internet as we know it. Until very recently, the prevailing sentiments were enthusiasm for the unprecedented ease of communication made possible by the spread of social media like Facebook, Twitter or LinkedIn; and euphoria for the unimaginable reach of powerful search engines like Google or Yahoo! This led hundreds of millions into writing off a "blind check" to the Internet, downplaying or outright ignoring its associated perils. Fuelled by massive network effects, as well as by a lenient regulatory framework which left them largely unregulated and not liable for user misconduct since the 1990s, Internet giants have risen at light speed: but they also brought with them risks, which they only partly anticipated, and for which they appeared far from well prepared, lacking both culture and ability to take mitigating actions. This was also evident when, after the Cambridge Analytica scandal, Mark Zuckerberg appeared before the U.S. Congress and the European Parliament to report on the fallacies of the largest social network in the world.

Ironically, the same features of the Internet that determined its rapid development also undermined its very foundations. Such features include the digital nature of the network, which makes it possible to copy and distribute files with no loss of quality and no dispossession of the original owners; and the end-to-end nature of the architecture, which makes it possible for every end user to communicate with every other end user. The combination of these effects led, over time, to an exponential increase in internet traffic: in 1992, 100 Gigabytes were exchanged between Internet users every day; by 1997, the same amount was exchanged per hour; in 2002, the same amount was exchanged per second. In 2016, the amount exchanged per second had reached 26,600 Gigabytes, poised to become 105,800 by 2021. In terms of annual global IP traffic, we are now living in the so-called "zettabyte age", with 1.2 Zettabytes being already reached in 2016, and expectations to reach 3.3 ZB by 2021. But this estimate might become even too conservative, if one considers the ongoing rise of the Internet of Things and the Internet of Value, alongside with the traditional Internet of Users[2].

The Internet today is however very different from its original version. Internet traffic is increasingly mobile, with smartphones expected to become a bigger source of traffic than personal computers before the end of the decade. Internet traffic is also increasingly made of content: globally, video accounts for three quarters of traffic. Cisco recently predicted that it would take "more than 5 million years to watch the amount of video that will cross global IP networks each month in 2021"[3]: by then, every second a million minutes of video content will cross the network. Moreover, Internet traffic is also increasingly non-human, with more than half of the traffic being generated by bots; increasingly hidden, with 88% of the whole Internet being inaccessible even to the most powerful search engines such as Google; and increasingly heterogeneous, with entire economic sector being gradually phagocytised by the network of networks, including financial services and energy.

---

[1] Quotes in: United States. Congress. Senate. Committee on the Budget (2012). *Concurrent Resolution on the Budget Fiscal Year 2013.* p. 95

[2] https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html

[3] Id.

This rising magnitude, and its associated complexity, led to important systemic consequences. First, the rise of an unprecedentedly information-rich society materialised the prophecy formulated decades ago by Herbert Simon (1955), who observed that "a wealth of information creates a poverty of attention". Navigating the clutter of the Internet, a gigantic copy machine unable to distinguish original content from artefacts and clones, required guidance and intermediaries. At the same time, competition for users' attention became the most prominent form of rivalry: those who would manage to conquer users' attention would get hold of the most precious resources in the Internet ecosystem. A resource that can be monetized in countless ways by selling advertising and a proliferating portfolio of services: thanks to direct and indirect network externalities, one-stop-shop effects led to the emergence of gigantic multi-sided platforms, able to grow at breath-taking pace by launching new services and reinvesting all profits; and ultimately leading to a consolidation of the dominion of a fistful of players, who keep competing "for eyeballs" by finding every day new ways to attract end users to their servers and portals[4]. The rise of platforms was at once an inevitable development, and a critical one, which made the Internet a most fragile giant.

Recent developments that affect the ability of end users to access quality content and applications, as well as the ability of large platforms to identify content and applications that are relevant for each end user, include the rise of Artificial Intelligence (AI). Currently limited to narrow applications such as optimisation of supply chains and extrapolation-based or correlation-based prediction of future events, AI (or, more appropriately, machine learning and training) increasingly plays a role in helping providers and users navigate the Internet thicket. As most enabling technologies, the evolution of AI can produce both good and bad outcomes, and normally a mix of the two: to say it with Stephen Hawking, effective AI could be "the biggest event in the history of our civilization. Or the worst". At the EU level, AI as recently the subject of a series of initiatives, which include the adoption of a Resolution by the European Parliament in 2016 (on civil law rules for robotics) as well as an ad hoc Communication by the European Commission published on April 25, 2018. But the potential for AI to affect the Internet ecosystem is still in its infancy, if not its embryonic level.

Against this background, the crisis has taken relatively long to emerge, but then has broken out in an unstoppable way, revealing all the vulnerability of the Internet. Scandals such as Cambridge Analytica and the alleged Russian meddling with U.S. presidential elections have pushed the accelerator on an already ongoing trend, particularly in Europe: the transition from a neutrality-based model of the Internet, in which intermediaries were not held liable for their user's conduct, towards a new model based on the responsible cooperation between public authorities and platforms, and the partial responsibility (or at least, the commitment) of large internet intermediaries for filtering out of their platforms undesired content such as hate speech, child pornography, elements of pro-terrorism guidance or proselytism; and yes, possibly also "fake news". Several recent initiatives of the EU institutions are shaping this trend, including codes of conducts for illegal content and even emergency interim measures on "digital taxation"; and the ongoing revision of the 2001 Information Society Directive and the 2000 e-Commerce Directive, true pillars and symbols of the early generation of EU Internet policy. The hearing of Facebook's CEO Mark Zuckerberg before the European Parliament on May 21, 2018 took place in this context, under the alleged failure to monitor and take action of the largest global social network. Lingering on the hearing will be the sense of displacement that US and

[4] See Wu (2017) The Attention Merchants. And Renda (2016).

EU policymakers perceive when dealing with these rising (or better, already risen) stars, and the impression that they "may have moved too fast and broken too many things"[5].

This paper, commissioned by the European Parliament, Committee on the Internal Market and Consumer Protection, analyses the consequences of these developments for online news and journalism, tackling the issue of so-called "fake news", or disinformation in the European Union. Section 2 below analyses the problem in detail, highlighting the main trends in news consumption and production, and the specific threats (if any) posed by the Internet age for what concerns the issue of disinformation. Section 3 explores the existing legal framework, starting from the trade-off between policies aimed at containing disinformation with the protection of fundamental rights such as freedom of expression; and encompassing also the recent report of the High-Level Group on disinformation created by the European Commission. Section 4 identifies a number of possible policy strategies that potentially address the issue of disinformation, ranging from self- and co-regulatory regimes to more prescriptive, command and control approaches. Section 5 briefly concludes with a number of policy implications.

[5] U.S. Republican Representative Greg Walden in addressing Mark Zuckerberg in the hearing before Congress on April 10, 2018.

2. DISINFORMATION IN MODERN COMMUNICATIONS

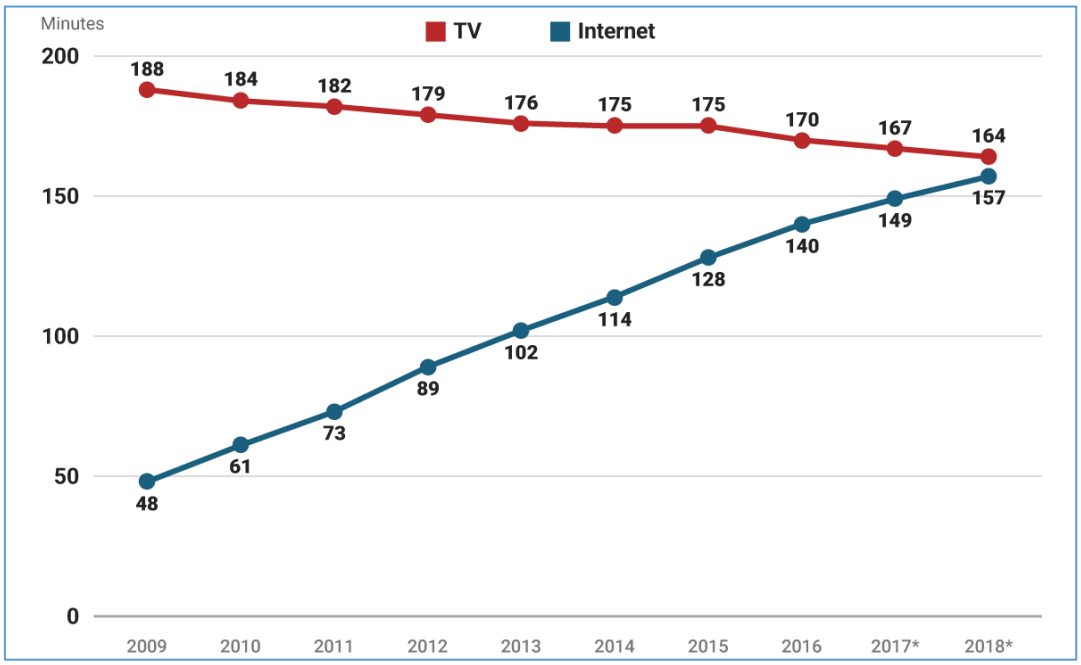
KEY FINDINGS

- Thanks to digital technologies, access to news has become easier and cheaper, and news consumption has grown unprecedentedly
- Online intermediaries have grown in a regulatory framework that made them not responsible for editorial control and to misconduct by their users. Online news are therefore often characterised by lack of fact- and originality-checking.
- Producing and spreading disinformation is much cheaper and easier thanks to the availability of a platformised, digital, end-to-end infrastructure for information exchange.
- The correlation between news outreach and quality-checks is broken. In the pre-Internet age, only media outlets with sufficient resources could reach out to large audiences. Today, anyone can be a publisher with global reach.
- The link between the content of the news and the means of communication is broken: as long as fake news bring traffic, money (i.a. from advertising), and no liability, there is no strong reason to expect that digital platforms will have a strong incentive to take independent action against disinformation.
- The quality of information on the Internet led to a variety of problems, including content bubbles, echo chambers, misinformation, disinformation and "online disinformation operations": the latter are the only ones that should be tackled by legislation on so-called "fake news"
- In today's information-rich world, no media and news channel is entirely shielded from possible political pressure or direction. Therefore, old and new media channels are more complementary than alternative.
- In an information-rich world, the most effective response to disinformation is (AI-powered) literacy.

2.1. Disinformation: yesterday and today

Give the rise of the Internet as a prominent means of communication, it is not surprising to note a massive increase in the online availability and consumption of news. As a matter of act, the global average daily consumption of news via the Internet is expected to reach consumption via television this year (see figure 1 below). According to Reuters, in the United States online news sources are already leading over television and other channels of news access.

Figure 1 Average media consumption per person worldwide, 2009-2018



Source: Zenith via Recode

* estimate

The spread of news via the Internet has both positive and negative implications. Among the former, **access to news has become easier and cheaper:** if one does not consider the exposure of end users to advertising as a form of payment, it would be possible to argue that most news have become free of charge. **News consumption has grown more than ever**, also thanks for the ongoing penetration of mobile connectivity and smartphone use at the global (including European) level. News are increasingly accessible, and increasingly easy to share exactly due to the phenomena described in the previous section: the digital, end-to-end nature of the Internet and its platformization with the emergence of (advertising-based) multi-sided business models[6].

This overflow of information, at the same time, caused problems. First, the **absence of responsibility and editorial control** for the quality of content flowing on internet platforms determined a **lack of quality and originality checks.** The almost unmanaged circulation of information on social media, mostly among peers, diluted the possibility to check the authenticity of news sources (Turk 2018). At the same time, the new form of news consumption has emerged **without a parallel development in media literacy:** over the past decade, the need to promote the enhancement of digital skills and the ability to discern original, authoritative content from fake or non-verified information has become a key concern of EU institutions. Once again, the promise of extreme democracy that came with the Internet **has fallen short of ensuring ways to distinguish good from bad content:** this is why search engines such as Google, based on an unprecedented (and possibly imperfect, but extremely helpful) way to rank content based on relevance and authority, became so **popular and at once critical on the Internet. Users need guidance on what to trust and what to discard, in order to avoid** getting lost in what philosopher Georg Wilhelm Friedrich Hegel would have defined as a "night in which all cows

[6] Clarke and Claffy (2015).

are black".[7] Similarly, Facebook's News Feed exploits the attention of Facebook users by exposing them to messages that have been subject until very recently to little or no filtering by the platform. This is at once inevitable, given the regulatory framework in which these platforms have evolved; and problematic, especially since in other policy domains, the **editorial control exerted by platforms such as both search engines and social media has been not only recognized, but indeed claimed** by the most prominent players[8].

To add to the overall complexity, it bears recalling that this phenomenon has produced a lot more than "fake news", whatever their definition[9]. As a matter of fact, the High-Level group set up by the European Commission to sharpen the definition of the phenomenon and discuss policy options recognized that the concept is elusive, and broadly refers to "forms of speech that fall outside already illegal forms of speech, notably defamation, hate speech, incitement to violence, etc. but can nonetheless be harmful". A closer look shows that the following related phenomena have emerged on the Internet:

- **Content bubbles (or echo chambers)** are described as a "state of intellectual isolation", which occurs whenever an individual interacts with a single news source, powered by an algorithm that only feeds users based on their perception of what they will like, or be interested in. Described in the past by Nicholas Negroponte and later by Cass Sunstein as "the daily me" problem, this problem is the product of both behavioural biases (such as the "confirmation bias", i.e. we tend to like what we already agree with)[10]; and the use of algorithms for personalized search, which are based on our past searches and thus mostly select content from a narrow subset of available sources. Well exemplified by the Wall Street Journal's "Blue Feed, Red Feed" site[11], the problem was officially acknowledged by Microsoft co-founder Bill Gates, who in a recent interview observed that the fact that on social media "you're not mixing and sharing and understanding other points of view" has turned out "to be more of a problem than I, or many others, would have expected".[12]The European Commission recently observed that "new technologies can be used, notably through social media, to disseminate disinformation on a scale and with speed and precision of targeting that is unprecedented, creating personalised information spheres and becoming powerful echo chambers for disinformation campaigns"[13].
- **Unintentional fakes/opinions (so-called "misinformation").** Anyone can express an opinion on the Internet, and share it widely on social networks. Even when one lacks a sufficient number of followers, there are strategies available (or even markets for followers, as on Instagram) that can maximise one's own reach in the social network communities. In this context, expressing one opinion, however false (e.g. "the Earth is flat"; "AI will create net employment"; or "trickle-down

---

7 Georg Wilhelm Friedrich Hegel, Preface to the Phenomenology of Spirit, translation and running commentary by Yirmiyahu Yovel, Princeton University Press, 2005, 248 pp, $19.95 (hbk), ISBN 0691120528.

8 See https://knightcolumbia.org/content/course-first-amendment-protects-google-and-facebook-and-its-not-close-question. And, for a recent and contrarian view, https://www.theguardian.com/commentisfree/2018/mar/09/google-facebook-first-amendment-protections.

9 See the Final Report of the High Level Group set up by the European Commission to advise on policy initiatives to counter fake news and disinformation spread online, at https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation.

10 Negroponte and Sunstein

11 http://graphics.wsj.com/blue-feed-red-feed/

12 http://www.bbc.com/news/world-us-canada-42187596

13 Communication "Tackling Online Disinformation", April 16 2018.

economics works for the poor"), can fuel disinformation even if there is no underlying intention to engage with manipulation of the public opinion. The Internet amplifies these statements, leading to an unprecedented rate of circulation of both true and false statements.

- **Intentional fakes or amplifiers ("Disinformation").** In a subset of cases, the lack of filtering on online platforms has led to the abuse of such platforms, with the clear intention to manipulate public opinion, for example in the occasion of an election. For example, news that Pope Francis had endorsed Donald Trump during the U.S. presidential campaign was spread by totally unreliable sources: but the overflow of information and the lack of attention of end users prevent many of them from spotting that sources such as "The American Patriot" or websites such as www.endingthefed.com were not exactly authoritative, or fact-based[14]. Many of these news are ignored by the public as clearly fake, but others spread very quickly and, even if at the margin, affect public opinion creating a thick layer of "noise" which causes trouble and confusion for end users.
- **Disinformation (or influence) operations.** These are commercially or politically motivated manipulation strategies, a variant of intentional disinformation, but happening at a much greater scale. These operations, which can be state-sponsored, aim at affecting the outcome of elections or at discrediting commercial rivals by purchasing privileged spots for online advertisements and using them to spread intentionally and strategically crafted messages. The Russian meddling in U.S. elections occurred exactly in this way: Facebook submitted a written statement to the U.S. congress, revealing that Russian agents created 129 events on the social media network during the 2016 U.S. election campaign: such events were viewed by 338,300 different Facebook accounts, 62,500 of which marked that they would attend. In 2016, campaign advertising on the internet skyrocketed in the U.S., increasing eight-fold since 2012 to an all-time high of $1.4 billion; and is projected to rise to $1.9 billion in the 2018 midterm elections, reaching 22% of all campaign ads. In reviewing its records, Facebook found approximately $100,000 in ad spending from June of 2015 to May of 2017 — associated with roughly 3,000 ads — that was connected to about 470 inauthentic accounts and Pages in violation of its policies; this led Facebook to infer that these accounts and Pages were affiliated with one another and likely operated out of Russia.

As appears evident, **not all these phenomena should lead to censorship or attempt to block the spread of information.** Divergent opinions always existed, and the diversity of opinions lies at the core of a sound political debate. Different opinions also often reflect different individual preferences, or even different cultural backgrounds. Asking internet intermediaries to filter out non-majoritarian, non-fact-based opinions from the major channels of access to information would dramatically impoverish our democracy and society; just like relying exclusively on neutral (relevance- or popularity-based) algorithms for the selection of news feeds would inevitably jeopardise the accessibility of local, niche or anyway "minoritarian" content (Renda 2015). The rise of AI-generated newsletters poses this problem, which requires a careful debate between policymakers and Internet intermediaries. I will come back to this issue in Section 4 below[15].

2.1.1. Objection: fake news always existed

Should we really worry about fake news as an unprecedented issue? After all, mis- and dis-information are old phenomena. Echikson (2018) reports that in the 13th century BC, Ramses the Great spread lies

[14] Presentation by the author at the European Parliament IMCO Committee hearing on fake news, 19 March 2018.

[15] https://www.wired.com/2015/10/this-news-writing-bot-is-now-free-for-everyone/

and propaganda portraying the Battle of Kadesh as a stunning victory for the Egyptians. The battle was actually a stalemate. And the Financial Times reports a fake news battle erupted when Julius Caesar appointed himself dictator for life in 44BC, a move that eventually led to his assassination that same year, on the Ides of March[16]. In modern times, in 1938 Orson Welles famously spread the word through a radio announcement that aliens from planet Mars had just disembarked in New Jersey: as many as one million radio listeners believed that a real Martian invasion was underway. History is obviously disseminated with egregious examples of misinformation and disinformation, notably including wars initiated, scientists burned alive, and massive witch-hunts organised due to mistaken beliefs and the spread of false opinions. So, one could say, why worry if there is nothing new?

As a matter of fact, there are good reasons to worry about disinformation in the Internet age, more than before. Part of these reasons have already been exposed. First, **producing and spreading disinformation is much cheaper and easier thanks to the availability of a platformized, digital, end-to-end infrastructure for information exchange.** Since disinformation operations are designed to capture the attention of end users by offering them tailored, catchy messages, it is no surprise that empirical evidence found a much faster spread of fake news compared to non-fakes. In a recent article appeared on Science magazine, Vosoughi et al. (2018) collected 12 years of data from Twitter and derived a set of 126,000 "fake news" stories shared on Twitter 4.5 million times by some 3 million people, concluding that a false story was 70% more likely to earn a retweet than verified news. While fake news was found in every category, false political stories were the most likely to be retweeted[17].

Second, the correlation between **news outreach and quality-checks is broken.** In the pre-Internet age, only media outlets with sufficient resources could reach out to large audiences. Today, anyone can be a publisher with global reach, with no need for sophisticated quality checking of news. This is important, since in the past the circulation of news was larger for those media outlets that could afford sophisticated quality check, which in turn would increase their reputation. This is not the case today: to the contrary, since news circulation has become essentially free of charge and mostly dependent on the installed user base of social networks, investing in quality checks has become uneconomical.

Third, another peculiar feature of this phenomenon is the **broken link between the content of the news and the means of communication.** Internet intermediaries, as already explained, have been traditionally shielded from responsibility for acts committed through, or on, their platforms. Absent specific legal rules in this sense, they remain private corporations with a profit motivation, which is directly linked to the amount of traffic they can generate, and the attention they can monetise in the form of advertising revenues. Rationally, as long as fake news bring traffic, money (i.e. from advertising), and no liability, there is no strong reason to expect that digital platforms will have a strong incentive to take independent action against disinformation.

Of course, things become different if civil society and political actors put pressure on internet intermediaries to take more responsibility for the content they host: this is what has been happening in the EU (and partly, also in the United States): over the past months, under the pressure exerted by **EU and national institutions in terms of moral suasion, law-making, and judicial decisions,** platforms such as Facebook and Google have started to engage in a much more granular oversight of the content that flows on their networks. In particular, the role of Facebook as content moderator and that of Google in enforcing copyright and the right to be forgotten are transforming these subjects into

[16] https://www.ft.com/content/aaf2bb08-dca2-11e6-86ac-f253db7791c6

[17] http://science.sciencemag.org/content/359/6380/1146

private regulators, or even more precisely, quasi-regulatory agencies in a hybrid, new form of co-regulation. This, also due to the rigidity of technology and the limited development of artificial intelligence, is also leading to inaccuracies in legal enforcement, often bordering on inadvertent forms of censorship (Echikson 2018).

These elements must be taken into account when crafting possible policy measures for contrasting disinformation at the EU level. While the Internet is evolving towards forms of co-regulation that involve enhanced responsibility of the digital platforms, this ongoing trend is also implying a relative inaccurate protection of key fundamental rights such as freedom of expression, often due to the rigid application of algorithmic protection of personal identity, personally identifiable data, or copyright. In the case of disinformation, as will be discussed below, the emerging trade-off between intervening to filter out alleged fake news and protecting freedom of expression deserves a high degree of caution when proposing policy measures.

2.1.2. Disinformation in an information-rich world: can we trust offline sources?

The disinformation debate is often presented in a way that demonizes online news sources in comparison to traditional news channels such as the printed press, radio or television. However, it is often neglected that, due to its more decentralised nature, the Internet can represent a key source of pluralism in contexts in which other forms of journalism are constrained and overly influenced by political powers. Notable examples include the role played by social media during the Arab spring, as well as in recent protests in Venezuela. During the latter protests, in the spring of 2014 political leaders of friendly neighbouring governments argued publicly that the photographs of civil unrest diffused on the Internet were indeed taken in Damascus, rather than Caracas. The rise of Donald Trump in the United States has been characterised by a trend towards "post-truth", with statements that come closer to René Magritte's ceci n'est pas une pipe than to an evidence-based reconstruction of reality. And even in the printed press, political agendas often lead to a dangerous walk on the tightrope between opinion-shaping and the illustration of verified facts: for example, the endorsement of Hillary Clinton as presidential candidate for the Democratic party by the New York Times in 2016 was criticised as purely political, and unfair towards Bernie Sanders; and the political polarisation of news channels such as Fox News is self-evident.

Against this background, media pluralism and the right to correct information require that citizens can have access a wide variety of sources, with different mixes of quality, political polarisation and control over the production process. In this respect, the following key pillars should be taken into account in the debate:

- **No media and news channel is entirely shielded from possible political pressure or direction:** the more we ask social networks to "polish" and "edit" content on their website, the greater the risk that content is steered towards a specific political direction. Note that this can happen: (i) due to **technical reasons** (political preferences of end users are fed as inputs to algorithms that then decide what is relevant and interesting for the end users themselves); due to **politically motivated decisions**, when the management of a social network decides to filter out or demote content that points at a specific political colour[18]; or (iii) due to **external manipulation:** this can occur both for search engines (See Renda 2015 for a taxonomy of techniques) and social networks (see Cambridge Analytica, but also Twitter's polluted population of fake accounts). In this respect, promoting

[18] For example, a controversial report by Gizmodo published in 2016 accused Facebook of systematically excluding conservative content. See https://gizmodo.com/former-facebook-workers-we-routinely-suppressed-conser-1775461006

pluralism also means mitigating the risk that a single source of news and information can capture the whole political debate.

- **Availability and exposure are two radically different concepts.** On the Internet, information is digital and over-abundant: the existence of a specific piece of information is almost never a problem. Rather, its exposure is much more important: as an example, past research has shown that approximately 50% of end users click on the first result of an algorithmic search on Google; whereas only 0.7% of end users click on the seventh. Likewise, all Internet intermediaries select and rank content in one way or another, mostly through algorithms: ending up in the twentieth page certainly means existence, but hardly means exposure.
- **There is more to pluralism than neutrality.** As I already observed in a past work (Renda 2015), imposing neutrality on Internet Intermediaries is a controversial, but possibly well-grounded concept at the infrastructure layer of the Internet ecosystem (in the so-called network neutrality debate); but it gradually loses meaning, the higher the layer of the internet ecosystem to which it is applied. Search engines and social networks that base their ranking and exposure of content on mere "objective" and "neutral" criteria will always pick the most popular, majoritarian content at the expense of local, niche content that is essential to provide end users with an alternative view.
- **Old and new media channels are more complementary than alternative.** Increasingly, especially if responsibility is placed on Tech giants, journalists will be hired as peer reviewers or even employees, with the key task of guaranteeing the quality of information exchanged and shared online. Journalists are also forming coalitions aimed at providing quality- and fact-checks of news that become viral. The Ethical Journalism network, FactCheck.org, TruthorFiction.com, Hoax-Slayer and many other services are emerging as new services that bring in the human dimension of journalists as an element of quality control of news that spread online. They will increasingly be able to use AI-enabled systems to detect news that spread more quickly and provide real-time checks, hopefully preserving human control over machines while performing their functions. Many academics have observed that new phenomena such as citizen journalism are more complementary than alternative to traditional mainstream journalism: however, this complementarity can only work if existing and future professional journalists will be rewarded for their work[19].
- **In an information-rich world, the most effective response to disinformation is (AI-powered) literacy.** Media literacy is essential for end users to be able to spot obvious fakes. At the same time, in the future human enhancement through AI-powered solutions may help us distinguish high quality news sources and content from bogus ones, and even spot potential political motivations behind the propagation of certain news. All this, however, requires dedicated political attention: AI, for example, is more likely to spread quickly on the supply side of Internet platforms than on the side of personal assistance to individual users, unless governments invest on this aspect of technology. And media literacy could be promoted both by governments and by Tech giants themselves, but this requires coordination and adequate monitoring to avoid that companies simply "whitewash" their operations by showing that they set up training sessions on how to read news.

[19] https://www.athensjournals.gr/media/2017-3-1-4-Noor.pdf

In conclusion, in an information-rich world all sources of information are needed, and complementary; skills and literacy has to be promoted and later enhanced through unbiased use of technology; media policy cannot limit itself to neutrality obligations, but must proactively seek exposure of a variety of quality news sources; and relying on the algorithmic take-down of content is very dangerous for freedom of expression, and unlikely to be proportionate to the overall goal of promoting the quality of information online.

3. POLICY RESPONSES: EUROPE AND THE WORLD

KEY FINDINGS

- •Many countries around the world, including many EU Member States, are taking action against disinformation. These laws often call on digital platforms to take action, but some also impose obligations on individuals.
- The possible balance between freedom of expression and the right to be properly informed lie in: (i) the promotion of responsible behaviour in conveying information to end users; (ii) the enactment of a proactive media policy aimed at promoting pluralism and improving the exposure of diverse content to end users; (iii) the empowerment of end users through media literacy initiatives, and supports to user behaviour.
- The European Commission has adopted initiatives in all of the three above-mentioned areas. The most important are the setting up of the High Level Group on Fake News and Online Disinformation, the Public Consultation on fake news and online disinformation and the Communication on "Tackling online disinformation". The latter announces the creation of a multi-stakeholder forum, a study to examine the applicability of EU rules and possible gaps, the creation of an independent European network of fact-checkers, the launch of a secure European online platform on disinformation, and the promotion of voluntary online systems allowing the identification of suppliers of information based on trustworthy electronic identification and authentication means, including verified pseudonyms, as provided under the Regulation on electronic identification.

The problem of online disinformation has been widely recognized around the world, especially after the alleged Russian meddling in the U.S. elections in 2016. Most of the associated legislative proposals are still pending approval, or have been approved very recently. A first glance at these initiatives enables us to build a first taxonomy of legal measures.

3.1. Traditional news process

Existing initiatives can be classified according to the following issues:

- **Scope:** what is defined as "fake news" or disinformation;
- Whether the **accountable** party is the Internet intermediary, or the original producer of the disinformation campaign;
- The **obligations** associated with accountability;
- The ultimate **remedies** and **sanctions** for non-compliance.

In the **United States**, the *Honest Ads Act*, a bipartisan bill, would require internet companies to disclose details on political advertisements placed on the companies' platforms. The proposed Act focuses in particular on foreign "meddling" into US politics by seeking to prevent "contributions, expenditures, and disbursements for electioneering communications by foreign nationals in the form of online advertising." Technology companies are troubled over the proposed Act, since it would compel companies to disclose details such as advertising spending, targeting strategies, buyers and funding. The have thus claimed that ongoing self-regulatory efforts can already tackle the problem

effectively[20]. For example, Twitter announced on 26 October 2017 — prior to the US Congressional hearings — its decision to ban Russian news outlets such as Russia Today (RT) from advertising on its platform.[21] Following the Senate hearings, the US government compelled RT to register with the Foreign Agents Registration Act (FARA) of 1938, which required individuals acting as agents of foreign influence with the capability to influence the government or public to "make periodic public disclosure of their relationship with the foreign principal, as well as activities, receipts and disbursements in support of those activities."[22]

In **Germany,** the *Network Enforcement Act* imposes fines as much as 50 million euros (US$53 million) on social media companies if they fail to remove "obviously illegal" content within 24 hours upon receiving a complaint. For offensive online material that requires further assessment, the Act compels companies to block the offending content within seven days, failing which a fine will be imposed. The Act mandated the establishment of a local point of contact for transnational technology companies to cooperate with local law enforcement authorities on takedown requests.

Some legislation proposals recommend tough penalties for individuals found responsible for disseminating false content. In the **Philippines,** for instance, the proposed Senate Bill No. 1492 threaten those guilty of creating or distributing fake news with a fine ranging from P100, 000 (US$1,950) to P5 million (US$97,587) and 1 to 5 years of imprisonment.[22] If the offender is a public official, fine and period of imprisonment will be doubled. Offenders will be disqualified from holding any public office.

In **Indonesia,** online smear campaigns had affected electoral candidates' standing in elections since 2012. There are evidence that some of these politically-motivated smear campaigns have been aided by well organised "fake news factories" such as the Saracen Cyber Team, an online syndicate that created many social media accounts to spread hate speech for clients willing to pay for them. Online sectarian narratives had polarised public opinion in the lead-up to the Jakarta gubernatorial elections in 2017 that saw the defeat of former governor, Basuki Tjahaja Purnama. The Indonesian government hence has beefed up existing legislations not only by introducing new provisions but also by issuing guidelines to aid their implementation and stepping up enforcement such as forming the Police Multimedia Bureau in 2017.

This Bureau may be similar to the Centre against Terrorism and Hybrid Threats in **Czech Republic, which also aims to counter disinformation campaigns.**

Some countries prefer to implement non-legislative measures such as fact-checking and counter fake news websites. **Malaysia** had introduced an information verification website (sebenarnya.my) to counter fake news while **Qatar** had launched the "Lift the Blockade" website to fight disinformation campaigns. Non-legislative measures may also include media literacy initiatives. Countries such as

---

[20] During the Senate hearings in November 2017, Facebook, Twitter and Google responded to questions on the role of technology companies during the 2016 election. Investigations revealed that Russian-linked entities such as the Internet Research Agency (IRA) used fake social media accounts to create content, which undermined the election process. Fake accounts were used to purchase ads and post politically divisive content in attempts to sow discord online. Facebook, for instance, has since estimated that Russian content had reached about 126 million Americans on its platform. Mike Isaac and Daisuke Wakabayashi, "Russian influence reached 126 million through Facebook alone," The New York Times, October 30, 2017, https://www.nytimes.com/2017/10/30/technology/facebook-google-russia.html.

[21] Dominic Rushe, "Twitter bans ads from RT and Sputnik over election interference," The Guardian, October 26, 2017.

[22] "Foreign Agents Registration Act (FARA)," The US Department of Justice, accessed November 28, 2017, https://www.fara.gov/.

Canada, Italy and **Taiwan** are introducing school curricula that teach children to discern between false and credible information.

Other recommended actions include regulatory measures such as identity management in registration of online domains. A legislative bill submitted to the **Italian** Senate in February 2017 require individuals who wish to open "an online platform aimed at publishing or disseminating information to the public" to notify the territorial tribunal via certified email, and provide the name of the platform, URL, name and surname of the administrator and tax number.

In **France,** a bill on disinformation was proposed in January 2018 and is still pending approval. The rules would be applicable only to online platforms which offer communication services professionally, for free or against payment, based on the classification or referencing by means of computer algorithms or on the linking of several parties for online content. These platforms are already subject to an obligation of transparency (article L 111-7 of the Consumer Code): the proposed Bill would reinforce this obligation by requiring that the platforms give clear, loyal and transparent information on who sponsored which content, and the amount received in exchange for sponsoring the content. Platforms will only be subject to this obligation for a limit time period: from the publication of the decree convening electors to the end of the voting process.

3.2. The legal basis in the EU

The legal basis for possible policy measures to counter disinformation operations finds a key constraint in the existence of a fundamental right to freedom of expression, which must be balanced with the right of the public to be properly informed.

Article 11.1 of the Charter of Fundamental Rights of the European Union (2000/C 364/01) recognizes the freedom of expression and information: "Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers." Similarly, **Article 10 of the European Convention on Human Rights** (hereinafter, ECHR) states: "1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises. 2. The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary."

The ECJ has reiterated the importance of this right and its application to all information and ideas[23]. Any limitations to the freedom of expression must thus be interpreted "restrictively"[24]. Further, any restriction "must be prescribed by legislative provisions which are worded with sufficient precision to

[23] Connolly / Commission (C-274/99), ¶ 39, citing ECtHR Handyside v United Kingdom judgment of 7 December 1976, Series A no. 24, § 49; Müller and Others judgment of 24 May 1988, Series A no. 133, § 33; and Vogt v Germany judgment of 26 September 1995, Series A no. 323, § 52.

[24] Id., at 41.

enable interested parties to regulate their conduct, taking, if need be, appropriate advice."[25] The European Court of Human Rights (hereinafter, ECtHR) has stated unequivocally that governments (and by extension the Union and the EU Review) cannot silence speech because it is "questioning the official view, being mindful that one of the main goals of freedom of expression was to protect minority views capable of contributing to a debate on questions of general interest which were not fully settled."[26] Even if content is false, the ECtHR has held that Article 10 ECHR "does not prohibit discussion or dissemination of information received even if it is strongly suspected that this information might not be truthful."[27]

The possible balance between freedom of expression and the right to be properly informed lie in:

- The promotion of responsible behaviour in conveying information to end users.
- The enactment of a proactive media policy aimed at promoting pluralism and improving the exposure of diverse content to end users.
- The empowerment of end users through media literacy initiatives, and supports to user behaviour, which can take the form of "hypernudges", i.e. ways to focus the attention of end users on a diverse set of contents, and to discourage them from sharing non-verified content.

The European Commission has taken initiatives that go in all three directions, as explained in Section 3.3 below.

3.3. Recent and upcoming initiatives by the European Commission

At the EU level, the final report of the **High Level Group on Fake News and Online Disinformation** recommended in March 2018 a "multi-dimensional" approach to the problem, at the same time refraining from advocating upfront regulatory measures[28]. Rather, the Report proposes a two-step approach starting with endorsed self-regulatory measures, and a subsequent evaluation already in spring 2019 to assess the need to move towards co-regulation, competition measures and/or monitored self-regulation during the next Commission term. The self-regulatory approach would go in parallel to interventions to strengthen media and information literacy and the diversity and sustainability of the digital information ecosystem, actions that by their own very nature have a longer time horizon. In addition, the Commission **Public Consultation on fake news and online disinformation** which took place between 13 November 2017 and 23 February 2018 confirmed that end users in the EU perceive this as a real problem: an analysis of the 2,986 replies (2,784 from individuals and 202 from legal organisations) showed that the problem is mostly related to social media, and that fact-checking by third parties is seen as the most effective potential remedy[29].

In April 2018, the European Commission echoed the findings of the High Level Group and those of a broad public consultation by adopting a **Communication on "Tackling online disinformation"**[30]. In the Communication, disinformation is defined as "verifiably false or misleading information that is

---

25 Id. at ¶ 42, citing Eur. Court H. R. Sunday Times v United Kingdom judgment of 26 April 1979, Series A no. 30, § 49.

26 Registrar of the European Court of Human Rights (ECtHR), Press Release: ECHR 370 (2013), EUR. COURT OF HUMAN RIGHTS (December 17, 2013).

27 Salov v. Ukraine, EUROPEAN COURT OF HUMAN RIGHTS (2005), ¶ 113.

28 https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation

29 https://ec.europa.eu/digital-single-market/en/news/public-consultation-fake-news-and-online-disinformation

30 https://ec.europa.eu/digital-single-market/en/news/communication-tackling-online-disinformation-european-approach

created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm". The latter comprises "threats to democratic political and policymaking processes as well as public goods such as the protection of EU citizens' health, the environment or security". Interestingly, the Communication distinguishes three phases of disinformation operations, including: (i) creation which includes recent, powerful forms of "deep fakes" which entail the use of false pictures and audio-visual content; (ii) Amplification through social and other online media, which can be algorithm-based, advertising-driven and/or technology-enabled; and (iii) Dissemination by users, who – as already mentioned – appear to be attracted by disinformation and more likely to share it compared to non-fakes.

Against this background, the Communication announces the following initiatives:

- A **multi-stakeholder forum on disinformation**, which will provide a framework for an efficient cooperation among relevant stakeholders and will develop an EU–wide Code of Practice on Disinformation to be published by July 2018, with a view to producing measurable effects already by October 2018;
- A **study to examine the applicability of EU rules and possible gaps in relation to the identification of online sponsored content**, which will also include the assessment of possible identification tools for online sponsored content.
- The **creation of an independent European network of fact-checkers** to establish common working methods, exchange best practices, achieve the broadest possible coverage across the EU, and participate in joint fact-checking and related activities.[31]
- The launch of a **secure European online platform on disinformation** to support the independent European network of fact-checkers, and enable them to act as trusted flaggers[32].
- The promotion, through the eIDAS Cooperation Network and in cooperation with platforms, of **voluntary online systems allowing the identification of suppliers of information** based on trustworthy electronic identification and authentication means, including verified pseudonyms, as provided under the Regulation on electronic identification.
- The promotion of research on ways to apply **new technologies such as artificial intelligence and blockchain to the issue of disinformation,** mostly through Horizon 2020 funds. A call will also be launched in 2018 for the production and dissemination of quality news content on EU affairs through data-driven news media.
- The launch of a **plethora of new initiatives** to raise awareness and promote media literacy and education.
- New initiatives involving the European External Action Service to **counter cyberattacks** involving disinformation operations.

The Commission added that it will report on the adoption of all these initiatives in December 2018. Unfortunately, the Communication was not backed by a formal Impact Assessment, aimed at evaluating the proportionality of these initiatives, which will entail the use of public funding in support

[31] The network will be invited to participate in the multistakeholder forum on disinformation.

[32] To this end, the Commission will consider the use of the Connecting Europe Facility and build on the experience gained in implementing the "Safer Internet" programme.

of a variety of policy interventions. At the same time, the approach adopted by the Commission appears sound, inasmuch as it does not venture into upfront regulatory proposals, but foresees an incremental approach, in which regulation kicks in only if the initial reliance on self-regulatory measures will prove insufficient or ineffective.

4. WHAT'S NEXT? POLICY OPTIONS FOR THE EU

KEY FINDINGS

- On the issue of cooperation with online platforms, simple transparency obligations with no additional commitments on the side of digital platforms would certainly be compatible with freedom of expression, but insufficient to tackle the phenomenon. At the other extreme, too rigid, command and control forms of regulation would likely be ineffective and disproportionate.
- A possible variant of transparency obligations, the creation of a black list of banned advertisers and media outlets, may be problematic and would fall short of providing appropriate incentives to online intermediaries.
- Monitored self-regulation and (if needed) co-regulation can enable important experiments such as: (i) the use of extensions for real-time fact-checking; and (ii) the use of "cybernudges" to induce end users to reflect before sharing. These options can of course be combined, and also integrated with additional obligations or commitments on the side of digital platforms.
- In a co-regulatory scheme the identification of non-verified news and possible fakes could be coupled with obligations to follow the money in detecting potential disinformation operations through advertising. Such option could also feature the use of performance or effectiveness indicators to enable monitoring and evaluation over time.
- EU media policy seems to be in strong need of developing co-regulatory solutions to increase the "exposure" of non-majoritarian content, and thus to proactively, rather than reactively, support pluralism.
- Media literacy programmes are essential to improve the resilience of EU citizens against the growing risk of disinformation. But they need to be made more comprehensive, including issues such as inter-cultural dialogue and appreciation for diversity of sources.
- EU policy must think long-term: online news are becoming increasingly an AI v AI battle: and the advent of deep fakes and generative adversarial networks may accelerate this trend.

The policy initiatives launched by the European Commission to counter the phenomenon of disinformation campaigns appears very meaningful and proportionate. However, an in-depth impact assessment would have contributed to clarifying the proportionality of the measures, as well as their compatibility with the legal basis as outlines in Section 3.2 above. Moreover, the real scope and governance of the proposed self-regulatory mechanisms, as well as the extent to which public authorities will be entitled to monitor compliance with established principles of conduct, remain to be fully clarified. Finally, of the three key areas outlined at the end of Section 3.2 above (the promotion of responsible behaviour in conveying information to end users, the enactment of a proactive media policy and the empowerment of end users), emphasis has been placed mostly on promoting responsibility on the side of online platforms; whereas a proactive media policy and an effective empowerment of the end users have remained in the background, with a general lack of constructive proposals.

Below, I outline a number of possible initiatives that could contribute to a more effective response to the identified policy problem, at the same time taking the peculiarity of the Internet into account.

4.1. Issue 1: Responsible and sustainable cooperation between platforms and public authorities

Within this first group of actions, several sub-options exist for the European Commission: I discuss them in an incremental way, from the least to the most prescriptive.

At one end, simple **transparency obligations** with no additional commitments on the side of digital platforms would certainly be compatible with freedom of expression. These obligations would include the provision to public authorities of information on the identity of advertisers, in particular of those that sponsored ads containing politically relevant or sensitive content, in line with what proposed in countries like France. This intervention could be coupled with enforcement measures directed towards the sponsoring entities, with no involvement of Internet intermediaries. At the same time, however, these measures may be insufficient to motivate digital platforms to take actions to contrast the phenomenon of disinformation operations, as well as to promote the quality of information exchanged on their platforms: problems that would most likely emerge in the implementation of this policy option include: (i) difficulties in identifying the types of messages for which information disclosure would be needed; (ii) possible confidentiality agreements between platforms and advertisers, which may lead the former to avoid disclosure in exchange for payments by the latter; (iii) difficulties in locating the original disinformation outlets, due to the use of mirror websites or other ways to hide the original sponsoring organization; (iv) extraterritoriality problems that limit the effectiveness of enforcement activities by EU public institutions. As a result, the simple introduction of transparency obligations may not be sufficient to tackle the problem and its underlying drivers.

A possible variant of transparency obligations would be the creation of **a black list of banned advertisers and media outlets.** This is the approach adopted, for example, by the East Stratcom unit set up in 2015 to rebut false and misleading stories about the EU, following Russia's hybrid war campaign in Ukraine. Based in Brussels, the unit has 14 staff and draws on volunteers and experts to monitor Russian media for the site the EU vs disinformation. The Disinformation Review carried out in this context seeks to control the right to freedom of expression by labelling publishers as "disinforming outlets" and their content as "disinformation," but was accused of possibly creating a chilling effect on the work of journalists that is central to democracy. Alemanno (2018) observes that "the labelling of publishers as 'disinformation outlets' is contrary to principle of the freedom of press established by the European Court of Human Rights". Apart from this, this option may face some of the constraints of the previous one: past experience, in particular on "follow the money" approaches in IP infringement, has produced mixed results[33] ; and the extreme volatility of IP addresses used as media outlets may lead to problems in locating and prosecuting operations, as has been the case throughout the history of the Internet.

Alternative options feature a more active involvement of digital platforms, based on the belief that they are best placed to mitigate the impact of disinformation operations. Among this family of options are **"monitored self-regulatory options"** based on a set of verifiable commitments. This option would have the advantage that is possibly stimulates innovation by allowing digital platforms to experiment

[33] See Batokas et al. (2018), Follow The Money: Online Piracy and Self-Regulation in the Advertising Industry, CESifo Working Paper No. 6852, at https://www.cesifo-group.de/DocDL/cesifo1_wp6852.pdf.

with alternative solutions to either filter out specific types of content (compatibly with freedom of expression), or empower end users through various forms of labelling, mandatory or optional fact-checking, and other potential solutions to the spread of disinformation.

More specifically, monitored self-regulation and more stringent regulatory modes such as co-regulation can enable important experiments such as:

- Use of extensions for real-time fact-checking. This would entail that a social network like Twitter or Facebook, or a search engine like Google or Bing incorporate in their news feeds and search queries the possibility of launching a real-time fact-checking powered by their own companies, by a selected third party, or even by roster of possible providers. Echoing the "ballot screen" Microsoft was forced to use a decade ago to enable user choice among competing browsers, it is conceivable that Twitter offers its users the possibility of choosing between a variety of fact-checkers, showing their average rating by end users. Such providers include companies like Poynter, the Ethical Journalism Network, PolitiFact.com, Snopes.com, and more. They typically employ a mixture of AI, data scientists and experienced human journalists. They could also be used as extensions on major browsers such as Chrome, Firefox, Internet Explorer and Safari.
- Use of "cybernudges" to induce end users to reflect before sharing. These could take various forms: for example, a social network could ensure that, whenever a piece of news is not considered as coming from a reliable source, a window would open in case the end user tries to share it, warning that this may constitute the spreading of fakes. This may increase the cost of sharing for the end user without leading to outright censorship of content: it would be a form of "libertarian paternalism", in which digital platforms nudge end users to a more responsible behaviour. A similar approach would be to introduce a "do not share non-verified news" option in the user settings of the digital platforms, and then enable the possibility for end users to change this feature by going in the user settings page. Empirically the reversal of default options (e.g. the use of two-side orienting as default option in office printers; or the selection of an opt-out mechanism for organ donation) has proven to be very powerful in affecting user behaviour.

These options can of course be combined, and also integrated with additional obligations or commitments on the side of digital platforms. For example, in a co-regulatory scheme the identification of non-verified news and possible fakes could be coupled with obligations to follow the money in detecting potential disinformation operations through advertising. Such option could also feature the use of performance or effectiveness indicators to enable monitoring and evaluation over time.

Compared to these monitored self-regulatory options, **co-regulatory options** are more stringent, in that they typically include more effective enforcement, such as sanctions for non-compliance with principles stated in legislation. They could, in any event, feature the same combination of conducts and policies outlined above, in particular the use of real-time fact-checking and the possible use of nudges. Enforcement options may then include, on the side of digital platforms and advertisers, settlement and arbitration procedures. And between platforms and end users, the activation of ex post fact checking in case a sufficient number of users request it. In addition, enforcement under co-regulation requires a legal backstop, which would enable enforcement by courts.

Finally, going beyond co-regulatory options to embrace more **command and control** regulation seems rather impractical at this stage, as it would lead to remedy that, in order to prove effective, would also be at once intrusive and disproportionate, leading to possible forms of inadvertent censorship. The German law that came into effect in January has led to what was considered by many as a for of

censorship, in that Internet Intermediaries have an incentive to take down all suspicious content with no nuance in assessing potential damage to free speech.

4.2. Issue 2: Towards a more proactive media policy

Apart from setting up a responsible and constructive cooperation with online platforms, a policy strategy to counter online mis- and dis-information (operations) would need to include the development of a more proactive media policy. The reason is that simply advocating more control of disinformation on the side of platforms would not solve the problem of polarization and lack of quality checks of news that are shared online. Moreover, simply stating that platforms have to be "neutral" or objective in displaying content may not be sufficient in order to promote and, where needed, restore adequate pluralism in news production and consumption. As a matter of fact, as already stated in Renda (2015), the notion of neutrality would lead most algorithms used by social media platforms and search engines to inevitably display in the most prominent spaces the most majoritarian content, rather than niche or local content. The notion of "relevance" of results and feeds, when it comes to media, may not be necessarily interpreted in favour of giving different views a chance to be seen.

To the contrary, EU media policy seems to be in strong need of developing co-regulatory solutions to increase the "exposure" of non-majoritarian content, and thus to proactively, rather than reactively, support pluralism. This is the first step towards an active policy to foster media literacy and a critical approach to both fake news and real, opinion news. A recent Recommendation of the Council of Europe (2018) also went towards this direction[34]. Recent academic papers such as Helberger et al. (2018) have also proposed to rely on exposure diversity as a core principle in the design of algorithmic recommender systems[35].

4.3. Issue 3: Empowering and supporting end users

The third pillar of a comprehensive strategy to counter online mis- and dis-information implies a focus on the end users, and their ability to distinguish reliable news from fakes, as well as distinguishing opinions from reality. This can be achieved in several ways, some of which are technology driven and thus less demanding on the end user; whereas others are based on media literacy initiatives and programmes, and as such require enhanced attention and competence on the side of the end user. More specifically:

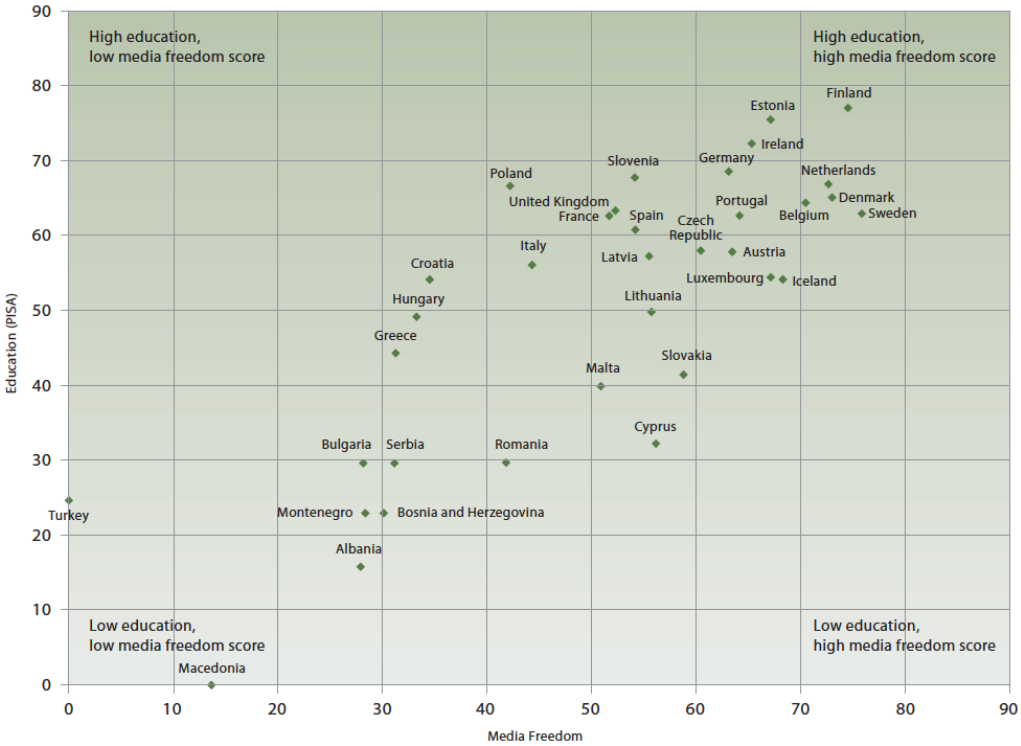
- **End users could receive a number of "signals" or additional services** from online platforms, or from a separate network of gatekeepers in charge of monitoring the quality and diversity of news flowing on online platforms. This could take the form, for example, of a network of "notaries" in charge of fact-checking and, more broadly, analysing news that get significant amounts of hits and are widely shared: such notaries could be public or private entities that share the burden of analysing news: end users may then be able to opt for receiving only "certified" news on their social media accounts.
- **Media literacy programmes** may be organised by social platforms themselves, as part of their corporate social responsibility; or by NGOs in cooperation with online platforms; or by universities or public institutions. Lessenski (2018) shows the correlation between media freedom and media education in 35 European countries.

[34] https://rm.coe.int/1680790e13 Recommendation (2018) 1

[35] https://www.tandfonline.com/doi/pdf/10.1080/1369118X.2016.1271900.

The graph shows the position of 35 countries in the index on two axes, based on their education (PISA) and media freedom scores, which are on a scale from 100 to 0 (highest to lowest)[36].

Figure 2 – Media Freedom and Education



A recent project that mapped existing media literacy programmes across the EU28 found that almost a third (305) of the main 939 media literacy stakeholders identified in the 28 EU countries were categorised as 'Civil Society' and all countries recorded main stakeholders coming from 'Civil Society. The next most common categories were 'Public Authorities' with 175 stakeholders and 'Academia' with 161 stakeholders allocated to these sectors. The study identified 189 networks, 135 of which were categorised as operating at a national level. Most of the project focused on skills such as 'Critical thinking' (403 of 547 projects) and media use (385 of 547 projects). The study also found that relatively few project tackled the issue of 'intercultural dialogue', which included skills around challenging radicalisation and hate speech online: this calls for more effort in future media literacy projects.

4.4. Longer term issues: towards "AI v AI" in journalism

Any policy strategy should think both short- and long-term. In the case of online disinformation, the increasing ability of operations campaigners to spread "deep fakes" and to hide behind obscure and impenetrable IP addresses calls for preparatory actions on the side of EU institutions. In particular, it is clear that the word of journalism will be permeated by artificial intelligence in the years to come. Current trends include i.e.: computational journalism and computer-assisted reporting; i-teams for algorithms and data; natural language generation for reading levels; computational photography; journalism as a service (JaaS), in which rather than reporting solely for their own publications,

[36] The education indicator uses PISA latest results, with reading having the highest weight (70%), and science and math with 15% each. The media freedom is based on the data from Freedom House and Reporters without Borders annual surveys, converted into standardized scores from 100 to 0 (highest to lowest).

journalists deliver content that can be used by third parties; real-time fact checking, synthetic datasets and more[37].

Deep fakes will become increasingly a problem. Political speech and even imagery can easily be manipulated at low cost and with professional quality: this, in turn, makes the transparency of the origin and circulation of content way more important as an element of fact checking. In a recent article appeared on Foreign Policy (2018)[38], Chris Meserole and Alina Polyakova reminded that "in the last two years, Kremlin-backed campaigns have spread false stories alleging that French President Emmanuel Macron was backed by the "gay lobby," fabricated a story of a Russian-German girl raped by Arab migrants, and spread a litany of conspiracy theories about the Catalan independence referendum, among other efforts"; and that in order to successfully tackle the problem, governments should keep an eye of emerging disruptive technologies such as **deep learning** and **generative adversarial networks** (GANs), which make it possible to manipulate images and video so well that it becomes difficult, if not impossible, to distinguish manipulated them from authentic ones. Apps like FakeApp and Lyrebird have made the production of "deep fakes" accessible to anyone.

Absent a real discussion on how to counter these development, any policy initiative on disinformation will become inevitable obsolete at birth. Accordingly, a constructive dialogue to the technology community will be needed to identify possible initiatives that may effectively contain or (less likely) eradicate the phenomenon.

[37] https://futuretodayinstitute.wetransfer.com/downloads/1e67e996768ac16ee87bcce1036f7ac020180301150134/c383da

[38] http://foreignpolicy.com/2018/05/25/disinformation-wars/

5. CONCLUSIONS

KEY FINDINGS

- The European Commission has proposed a set of meaningful measures: the key is to set up a constructive dialogue with online platforms, mobilise and coordinate fact-checkers, and promote media literacy.
- The incremental approach to regulation (first self-, then co-regulation if needed) proposed by the Commission is meaningful: however, self-regulation should be accurately monitored through the definition of indicators, and the sharing of practices, which can stimulate innovation.
- There is a need to avoid rigid solutions that would amount to censorship. Command and control regulation cannot achieve meaningful results in this field.
- The future holds uncertain developments for online news: research on AI systems must be carried out in an integrated way with online intermediaries, other tech companies, and cybersecurity experts.

This paper has surveyed the key outstanding issues in Europe in the domain of online disinformation. In a world characterised by decreasing trust in the Internet and increased inability of end users to single out news that are reliable and worthy of their attention, the role of online platforms is crucial and critical. Crucial, since no traditional form of law-making can succeed without cooperation with platforms, and thus either self- or co-regulation schemes. Critical, since online platforms may not, absent pressure from public policy, have enough incentives to take action to filter news that bring money, attention, and traffic to their sites.

The series of measures contemplated in this paper, thus, start and focus in particular on monitored self-regulatory measures that involve online platforms as gatekeepers. These measures have the added value of being flexible (as the subject matter constantly changes), and accommodating behavioural solutions such as nudges, which can empower end users and de-bias their behaviour in approaching news consumption.

In addition, this paper reaches the conclusion that the proposed measures would not be sufficient to promote a more sustainable evolution of the online news market: the policy mix (as largely anticipated by the European Commission) would need to include a proactive media policy oriented towards "exposure" and pluralism; an ambitious plan to promote media literacy and user empowerment; and a forward-looking approach to innovative solutions for the era of "deep fakes".

REFERENCES

- Alemanno, A., J. Brogi, M. Fischer-Zernin and P. Morrow (2018), Is the EU Disinformation Review Compliant with EU Law? Complaint to the European Ombudsman About the EU Anti-Fake News Initiative (March 28, 2018). HEC Paris Research Paper No. LAW-2018-1273. Available at SSRN: https://ssrn.com/abstract=3151424 or http://dx.doi.org/10.2139/ssrn.3151424
- Batokas et al. (2018), Follow The Money: Online Piracy and Self-Regulation in the Advertising Industry, CESifo Working Paper No. 6852, at https://www.cesifo-group.de/DocDL/cesifo1_wp6852.pdf.
- Cisco, Visual Networking Index (2018), https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html
- Claffy, K.C. and D. D. Clark (2013), "Platform Models for Sustainable Internet Regulation", TPRC 41: The 41st Research Conference on Communication, Information and Internet Policy, 15 August (http://dx.doi.org/10.2139/ssrn.2242600).
- Clark, D.D. and K.C. Claffy (2015), "Anchoring policy development around stable points: an approach to regulating the co-evolving ICT ecosystem", Telecommunications Policy, in press, August 2015.
- Echikson, W. (2018), Privatising Censorship, CEPS 2018.
  At https://www.ceps.eu/publications/privatising-censorship
- Gawer, A. (2009), Platforms, Markets and Innovation, Edward Elgar, Cheltenham, UK and Northampton, MA.
- Gawer, A. and M.A. Cusumano (2008), "How companies become platform leaders", MIT/Sloan Management Review 49 (2), 18–35.
- Helberger, N., K. Karppinen & L. D'Acunto (2018) Exposure diversity as a design principle for recommender systems, Information, Communication & Society, 21:2, 191-207, DOI: 10.1080/1369118X.2016.1271900
- Martens, B.; L. Aguiar, E. Gomez-Herrera and F. Mueller-Langer (2018), The digital transformation of news media and the rise of disinformation and fake news - An economic perspective; Digital Economy Working Paper 2018-02; JRC Technical Reports.
- Renda, A. (2016), Regulation in the layered Internet ecosystem: challenges and myths, in the Book "A Level Playing Field for the Digital Ecosystem: Protecting User's Rights Online, written by the Technical University of Madrid and Commissioned by Fundación Telefónica.
- Renda, A. (2016), Selecting and Designing European ICT Innovation Policies, Report for the European Commission, Joint Research Centre and Institute for Prospective Technological Studies, forthcoming April 2016.
- Simon, H.A. (1956), "Rational Choice and the Structure of the Environment", Psychological Review, Vol. 63, No. 2, pp. 129-138.
- United States. Congress. Senate. Committee on the Budget (2012). Concurrent Resolution on the Budget Fiscal Year 2013. p. 95
- Wu, T. (2017) The Attention Merchants. Atlantic Books.

Contact: Poldep-Economy-Science@ep.europa.eu
This document is available on the internet at: www.europarl.europa.eu/supporting-analyses

PE 619.013
IP/A/IMCO/2018-04

Print ISBN 978-92-846-3149-0| doi: 10.2861/685258| QA-03-18-042-EN-C
PDF ISBN 978-92-846-3150-6| doi: 10.2861/468200|QA-03-18-042-EN-N