Alliance of Democracies

Disinformation, Deepfakes & Democracy

The European response to election interference in the digital age

#DefendDemocracy

About the publication & author

This report has been written by Christoffer Waldemarsson and published by The Alliance of Democracies Foundation on April 27th 2020 as a contribution to the continuous work of the Foundation, including the activities of the Transatlantic Commission on Election Integrity and its mission to highlight the issue of foreign election interference and protect the integrity of democratic elections. The report and its recommendations reflect the opinions of the author.

Christoffer Waldemarsson holds a Bachelor of Science in Economics and Political Science from Lund University and a Master of Science in International Business and Politics from Copenhagen Business School. With a background in the Swedish Ministry for Foreign Affairs and the Nordic Council of Ministers, he has previously published reports on the state of democracy in the Nordic countries and academic articles on what effect democracy has on economic growth. Christoffer Waldemarsson is working at the Alliance of Democracies Secretariat in Copenhagen.

The Alliance of Democracies Foundation

The Alliance of Democracies Foundation is a non-profit organization founded in 2017 by former Danish Prime Minister and NATO Secretary General Anders Fogh Rasmussen, dedicated to the advancement of democracy and free markets across the globe. With a rising scepticism towards free and liberal values and a retreat in American global leadership, the foundations aims to become the world's leading "megaphone" for the cause of democracy and to encourage greater cooperation between nations who share common values, to counter spoilers and support those who wish to bring more democracy to their countries. More information about the Foundation and its activities can be found on allianceofdemocracies.org.

The Alliance of Democracies has a number of ongoing initiatives focusing on the democratic challenges stemming from new technological advancements to ensure that tech is used for benign purposes, among them the Campaign for Democracy. The main initiative in the Campaign for Democracy is the Transatlantic Commission on Election Integrity, a group of leaders in politics, tech, media and business that have united across the Atlantic and transcending party lines to help prevent the next wave of foreign election interference. The Commission seek to raise awareness in the public and of governments on what foreign powers are doing to undermine liberal democracies, and aim to recommend concrete actions to governments and tech companies.

Copenhagen · Denmark · info@allianceofdemocracies.org

Table of Contents

Introduction 5

Digital disinformation and social media 6

Deepfakes – the new wave of disinformation 8

Implications on democracy and society 10

What can be done? 11

Voluntary restrictions by platforms 11

Towards regulation – the free speech controversy 13

The EU response to disinformation 14

The European Union 15

Germany 17

France 18

The United Kingdom 19

Recommendations and lessons learned from the EU 21

Introduction

Foreign election interference is by no means a new phenomenon, and the use of disinformation to change opinions of citizens or sway the outcome of elections is a strategy at least as old as the printing press.[1] Recent years has seen an exponential increase of attention directed towards foreign disinformation campaigns however, primarily spurred by the Russian intervention in the 2016 US presidential election and the reports of foreign interference in the Brexit referendum that same year. And these prominent examples are just the tip of the iceberg. Reports of Chinese disinformation campaigns to discredit the China-critical presidential candidate in the Taiwanese election in 2020[2], suspicion of Russian attempts to sway the UK election in 2019[3] and proof of Chinese fake accounts campaigning against pro-democracy activist in Hong Kong in 2019[4] are other recent instances where this type of digital election meddling has reached the front pages of international media – not to mention the false information and online rumours surrounding the Covid-19 outbreak in 2020. Disinformation, defined as false information spread to intentionally deceive the public, seems to have become the new "normal" and a recent study suggested that "the majority of people will see more false than true information in 2022".[5]

The increased prevalence and efficiency of disinformation campaigns can be linked to technological advancements and the changing media environment in our society. With information travelling faster than ever and new actors responsible for the distribution of news we can rightly talk about *the new information age*, where the old model for gatekeeping of content carried out by journalists and publishers has been disrupted.[6] The introduction of social media and the algorithms these platforms are built on enables disinformation to spread at an unprecedented speed, making a coordinated attempt to influence the perceptions and opinions of citizens more dangerous than ever before. The social media platforms are also constructed in a way that encourages provoking content and material shared by people with similar opinions, tying to a number of human biases that makes users of social media even more susceptible to the information appearing in their feed.[7]

As free and unbiased elections are a pivotal part of our democratic societies, dependent on the ability of citizens to make informed decisions based on accurate information, the need to address the increased use of disinformation campaigns by foreign states poses a crucial challenge for our democracies. When voters are faced with false narratives and information it destabilizes their sense of certainty and causes confusion,[8] with the potential risk of having a direct effect on a democratic election by changing the outcome. But disinformation may also have an indirect effect on democracies, as it can increase social division in society, lower trust in institutions and risks to weaken democratic discourse on a societal level.[9] Disinformation campaigns also often implicitly support extremist ideas and, in the case of Europe, carry an EU-critical tone making it a threat to both legitimacy and unity among the member states.[10]

[1] (European Parliament, 2020)

[2] (Aspinwall, 2020)

[3] (Scott, 2019)

[4] (Palmer, 2019)

[5] (Scheidt, 2019)

[6] (Chesney & Citron, 2019)

[7] Ibid.

[8] (Scheidt, 2019)

[9] (Jaursch, 2019)

[10] (Scheidt, 2019)

A report published by the Freedom House examining elections in 2016 found that online manipulation and disinformation tactics played an important role in at least 18 national elections that year, including the United States,[11] and the Transatlantic Commission on Election Integrity has seen an increased activity of automated Twitter accounts before national elections in their election monitoring. The upcoming US presidential election in 2020 and the anticipated efforts to manipulate media and influence the result of the electoral process across the Atlantic emphasizes the urgent need for concrete action, but the problem is just as relevant in a European context. The previously mentioned Brexit referendums, the "Macron leaks" in the 2017 French presidential election and reported concerns about foreign disinformation attacks before the 2019 European elections are examples of the increased prevalence of the issue also in the European Union.

The EU and its member states were slow in their initial response[12] but have recently stepped up their efforts to tackle this new threat to our democratic processes, with measures ranging from strict regulatory demands to public awareness campaigns and voluntary restrictions for the social media companies. The social media platforms have also taken their own initiatives to curb the spreading of disinformation. While it may seem tempting to enforce hard legislation, regulatory measures are subject to criticism when trying to balance efficient ways to tackle disinformation and the risk of hindering free speech and restrict the open public debate,[13] a part of our democratic societies just as central and necessary as the access to accurate information. This report will address the increasingly prominent threat of digital election interference by first explaining the role of social media and the new technological phenomenon *deepfakes*, with its potential effect on democratic states. Thereafter, the report will map out the voluntary restrictions taken by the social media companies themselves, as well as the measures taken by a selection of European countries to mitigate the spread of disinformation. The report will then end with some concluding recommendations and lessons learned from the European countries.

Digital disinformation and social media

While there is no doubt that rumours and falsehoods have appeared also in a historical perspective, especially in countries where media is heavily partisan in nature,[14] cyberspace and the digitalization of electoral processes have surely taken disinformation to a new dimension. Disinformation campaigns before the introduction of internet were mostly limited to printed media and the written word but can now be spread from a number of devices and platforms, while also enabling easier manipulation of pictures and video material.[15] Citizens are constantly connected and information travels across continents at an unprecedented speed. Social media platforms are a large part of the new information and communications environment - and an important factor taking disinformation campaigns beyond its previous scope and effectiveness as they have changed the way in which news and information is both produced and consumed.

[11] (Freedom House, 2017)

[12] (Scheidt, 2019)

[13] (Schmitz-Berndt, M. Berndt, Rechtsanwalt, & Rechtsanwälte)

[14] (Kofi Annan Foundation, 2020)

[15] (Scheidt, 2019)

A study of the 2016 US presidential election found that Twitter users in the US shared as much "junk news" lacking professional journalism as they shared professionally produced news in the time leading up to the election,[16] with many of these 'fake news' stories outperforming professional news. While arguments framing the issue with disinformation campaigns as purely a "social media challenge" largely ignores the role traditional media has played in spreading disinformation[17], there are a number of factors that could explain *why* social media and the new ICT's plays such an important role in the spreading of disinformation.

(I) Disrupt old distribution models and content gatekeeping

Thirty years ago, the practical ability of individuals and organizations to spread and distribute articles, images, audio or video was limited – irrespective of the authenticity of the content. Instead, people relied on news agencies, media organizations and the newspapers you had subscribed to.[18] For good and for bad, the number of outlets providing information in society were few and, at least in liberal democracies, under limited influence from government and external actors with self-interest. The information revolution has disrupted this content distribution model, democratizing access to communication and lowering the quality control of what information reaches large groups of people.[19] A substantial part of the gatekeeping previously overseen by editors and publishers has now shifted to the digital platforms, where almost anyone with a keyboard now instantly can publish their material to potentially large audiences.[20]

Another aspect of the disrupted distribution models is that the previously dominating traditional media, in desperate attempts to retain and perhaps regain the attention of readers, now may be *incentivized* to spread doubtful content in their own channels. If a news story is viral enough, it is easy to imagine traditional media prioritizing fast distribution over thorough background checks of the source material. And even if the background research is properly executed, the sheer fact that a news story has gone viral may be a valid reason to publish a story of doubtful quality –increasing the reach and spread of content published in social media channels even further.

(II) Reward negative content and extremist voices

At the core of Facebook, Google, Instagram, Twitter and YouTube, just to name a few examples of new and popular media channels, lies the logic of maximizing attention – a logic for which anger and lies often work better than objectivity and truth.[21] The algorithms these platforms are built on are not neutral and can intentionally promote inflammatory, or even false, content to their users in order to keep their attention for the longest time possible and subsequently maximize advertising revenues.[22] [23] This ties to a human bias - *our natural tendency to engage with negative and novel information.* Empirical evidence examining 126,000 news stories on Twitter between 2006 and 2010 show that hoaxes and false rumours reached people *ten times* faster than accurate stories, and that falsehoods were 70% more likely to get retweeted than accurate news.[24] The researchers found that this was not due to bots spreading the stories, but rather because *people* shared inaccurate news items

[16] (Ibid.)

[17] (Kofi Annan Foundation, 2020)

[18] (Chesney & Citron, 2019)

[19] Ibid.

[20] (Kofi Annan Foundation, 2020)

[21] (Jaursch, 2019)

[22] (Jaursch, Lenoir, Schafer, & Soula, 2019)

[23] (Kofi Annan Foundation, 2020)

[24] (Vosoughi, 2018)

more frequently. The effectiveness of disinformation distribution is therefore dependent on how these algorithms are designed and the policies behind these mechanisms are central in understanding the spread of false content.

(III) Create segmented, personalized feeds and information bubbles

While the introduction of social media has democratized the political and societal debate, enabling more people to participate, it is also easier than ever for citizens to cement their already pre-existing beliefs about society. As platform algorithms highlight popular content shared by friends, and because people tend to share information with which they agree, social media users are surrounded by information confirming their pre-existing views.[25] This ties to a human bias called *the filter bubble*, where we naturally tend to surround ourselves with information confirming our beliefs.[26] It also relates to the increase in the sheer volume of information people are presented with today, making it challenging to distil true facts from false claims. False or misleading information on key political topics can quickly "go viral" online, and part of the reason is another human bias, *information cascade*. When the amount of information has increased, people stop paying attention to each and every piece of content and instead rely on what they assume other have reliably assessed and chosen to share.[27]

Microtargeting is an example of how these filter bubbles are used, where commercial and political organizations can direct their message to audiences specifically prone to be affected and incited by their content.[28] Personal data can be used to create detailed profiles that enable companies to sell targeted ads, further amplifying the effect of information campaigns. Without the gatekeeping of traditional media, deceiving people and manipulating societal debates is easier than ever.[29] These types of targeted ads are part of the social media platforms business model. They make money as marketing platforms by selling personalized ad spaces, without an obligation to educate users or a journalist's ethical restraints on what to publish. The longer the companies can keep their users on their websites, the better it is from a business perspective – enhancing the effect of personalized feeds. Combining all of these factors, including the human biases and a general growth in scepticism towards political institutions and media, it is easy to see how the technological advancements enabling rapid sharing of information and the introduction of social media platforms have increased the effectiveness and threat a coordinated disinformation campaigns poses to democratic states and electoral processes.

Deepfakes – the new wave of disinformation

At a time where it is already increasingly difficult to tell facts from fiction and information from disinformation, technological advances make it difficult to trust even what we see with our own eyes. The introduction of deepfake-technology, a way to digitally fabricate video and audio using AI and machine learning algorithms, enables the insertion of faces and voices into video recordings of actual people to create realistic impersonations. We are seeing increasingly realistic and convincing

[25] (Chesney & Citron, 2019)

[26] (Ibid.)

[27] (Chesney & Citron, 2019)

[28] (Jaursch, 2019)

[29] Ibid.

manipulated videos of prominent politicians like Barack Obama and Donald Trump say and do things taken out of thin air, and with plenty of footage available on the person being depicted, the impersonations become increasingly accurate.

The technology to create deepfakes often involve the use of "neural networks", a set of algorithms designed to recognize patterns, for machine learning and it is when these networks processes a broad array of training examples that increasingly accurate models can be created.[30] The technology is further advanced by the use of *generative adversarial networks* (GANs), which brings two neural networks to bear at the same time. In an iterative process, these two networks simultaneously produce and assess a sample from the dataset with a speed, scale and nuance that human reviewers can't achieve. Leaving the majority of the technical specifications that goes into making this type of fabricated material aside, the rapid developments in the field undoubtedly enables more sophisticated fake videos, but also makes it more accessible for people with less developed technological skills. The diffusion of the technology can be seen by just googling "deepfakes" or make a simple search on video platforms such as YouTube. Private individuals are able to produce content depicting famous actors, politicians and other public figures with realism that is more or less impossible to distinguish from reality. New Deepfake-apps are also popping up frequently, like Chinese app Zao, allowing you to swap faces with well-known actors in movies and TV-series,[31] or the desktop app Fake App, which enables "photorealistic faceswap videos by deep learning",[32] again exemplifying how the technology is being democratized. And there is no reason to believe that trend is turning.

By the US presidential election in 2020, some experts expect these tools to have become so widespread that anyone with a little technical knowledge will, from the comfort of their home, be able to make a video of any person doing and saying whatever they want.[33] With deepfakes often defined as manipulated material and product of artificial intelligence and machine learning, it can be separated from other types of manipulated media such as "*cheap fakes*" and "*shallow fakes*". The two latter terms are closely related to deepfakes but refers to content that is just edited or cut in ways to change the material without the use of AI[34] – a much simpler way of manipulating for example a video clip. *Cheap fakes* and *shallow fakes* could still serve the same effect as deepfakes, however, not least seen in the clip from 2019 where Nancy Pelosi's speech had been slowed-down and appeared suspect.

While the main focus of this paper will be on the negative implications deepfakes may induce on democratic states, it should be mentioned that the technology by now means need to be used with illicit intentions in mind. Deepfake technology can create several new opportunities in the field of education, medicine, arts and cinema, just to name a few examples.[35][36] In addition, we have seen a recent example of how deepfakes can be used for "positive campaigning" in an election, as a political leader in the 2020 Legislative Assembly elections in Delhi manipulated a video to reach new linguistic groups in his constituency.[37] With that in mind, the rest of the report will predominantly

[30] (Ibid.)

[31] (Murphy & Huang, 2019)

[32] (Chesney & Citron, 2019)

[33] (Chertoff & Donahoe, 2018)

[34] (Edelman, 2020)

[35] (Chesney & Citron, 2019)

[36] (Chandler, 2020)

[37] (Nilesh, 2020)

focus on the negative implications deepfakes may have on elections, with the most important effects listed in the next section. These points are just as relevant in a broader discussion of disinformation, as deepfakes can be seen as a representation of the latest advancement of previous methods – and if used efficiently a very dangerous advancement.

Implications on democracy and society

(I) Manipulation of elections

The perhaps most apparent way in which deepfakes can influence a democratic process is the direct use of false material to change an election result. Imagine a video where a presidential candidate or prominent figure in a political party acts inappropriately or says something controversial. How would the voters react? With small electoral margins and rapid communication, it is no stretch imagining how such a happening could have drastic effects on an election outcome. This type of content is especially dangerous if the distribution of the material is timed early enough for it to go viral, but without enough time to debunk it before an election.[38] Foreign states with an interest in the election would be the obvious candidate for such an attack, but with the lowered barriers to using deepfake technology, the risk is that any individual with basic technical abilities could coordinate a similar attack as soon as by 2020.

(II) Exacerbating social divisions

In the polarized political climate of today, deepfakes and disinformation could severely worsen pre-existing social divisions. Releasing manipulated video material that confirms (or seems to confirm) polarizing arguments based on economic inequalities, race, sexuality or ethnicity could not only lead to a stronger division, but in worst case also accomplish mobilization and call for action that the written word could not.[39] Countries where polarization is already strong, with pre-existing scepticism towards media and a tradition of partisan coverage are much more vulnerable than countries with higher trust.[40] Releasing a deepfake that exacerbates social divisions in society may have both short- and long-term effects. In the short term, it could have an indirect implication for an election result, benefitting parties with a clear stance on issues highlighted in the material released. In the long term, it could lead to a general *"us-against-them"* attitude in society, with declining trust in fellow citizens, becoming a breeding ground for populist opinions and democratic backsliding.

(III) Lowering trust in institutions and authorities

Just as in the previous point, deepfakes may also have a more indirect effect on elections and democracy as they affect citizen's trust in institutions and public authorities. A fake-but-viral video of a police officer acting violently, a judge privately discussing ways to circumvent the judiciary system or border guards using racist language could all have devastating effects on the trust in authorities, and in the current climate perhaps also lead to action from citizens. Deepfakes targeting public authorities also spur on the currently broader trend of more distrust in society and is especially dangerous when citizens start to question information that comes from their own public institutions, with risks for both democracy and public safety.[41]

[38] (Chesney & Citron, 2019)

[39] (Ibid.)

[40] (Kofi Annan Foundation, 2020)

[41] (Chesney & Citron, 2019)

(IV) Undermining journalism and information

A gradual decline in the media's trust can already be seen on a global scale[42] and deepfakes and disinformation will do little to help this situation. As it becomes increasingly difficult to tell reality from fake, the trust in media is under even more pressure. Not only are the requirements to rapidly debunk or judge the authenticity of material higher than before, the public can now also point to every news story that is not confirming their view on society and claim it to be fake – and back it up with examples of when news organizations have made mistakes in the past. How can the authenticity of material be guaranteed? There will undoubtedly be more cases of news stories stemming from material that will later be deemed false, and this might even lead journalistic outlets to hesitate in publishing news stories altogether.

Another important aspect of the harms deepfakes can induce is the scale and ease of debunking. While a Deepfake depicting a presidential candidate may have a large direct effect and headline every newspaper, the enormous interests in such an event would also mean that a lot of resources will be put into debunking it. The type of deepfakes that flies under the radar, for example showing violent behaviour of a police officer in a small region or state, some specific religious groups acting inappropriately or the unlawful handling of a peaceful protester by public authorities may prove more dangerous in the long term. Deepfakes that are not viral enough to be immediately debunked but still affect the opinions of the public are dangerous just because they do not become headlines.

What can be done?

Voluntary restrictions by platforms

With many governments being slow in their initial response to digital election interference, the perhaps most comprehensive efforts to stop the spread of disinformation to date has been taken by the tech companies themselves. Although self-defined restrictions differ by design and target, there is no doubt that these voluntary restrictions on content imposed by social media platforms play an important role in mitigating the spread of disinformation. The ground on which the voluntary content moderation relies stems from their rights as publishers, determined by terms-of-service agreements. The terms-of-service agreements between users and social media platforms are perhaps the most important documents governing digital speech in today's world.[43] Generally, the measures can be divided into two groups, either targeting the *exposure* of disinformation, often in the form of content removal, or targeting users' *beliefs in false content*, which usually rather takes the form of labelling.[44]

Twitter announced in November 2019 that they would ban all political ads to prevent the spread of disinformation in election campaigns.[45] This was followed by an announcement in February 2020 imposing a partial ban on manipulated visual material, in practice meaning deepfakes. All published deepfakes and doctored images that may result in "confusion or misunderstanding" on Twitter will from March 2020 and onwards be labelled "manipulated", and users intending to

[42] (Kofi Annan Foundation, 2020)

[43] (Chesney & Citron, 2019)

[44] (Metzger, 2019)

[45] (BBC, 2019)

share such material will be shown a warning.[46] YouTube, a platform owned by Google, has also taken action and announced just before the first 2020 US primary election in Iowa that they will reinforce their guidelines on misleading election-related material. This includes the removal of any content that has been "technically doctored", manipulated or aims to mislead voters.[47] An initiative to improve the identification of deepfakes on the internet has been taken by Microsoft and Facebook among others, who have launched the Deepfake Detection Challenge, inviting people around the world to build innovative new technologies that can help detect deepfakes and manipulated media.[48] The initiative is set-up as a competition running in spring 2020, aiming to develop the best possible tools in detecting AI generated videos.

Facebook has come under heavy scrutiny in recent years over their responsibility in tackling disinformation. The social media platform has made a point out of not banning political ads with reference to the free speech protection in the US Constitution's First Amendment, stating that they do not want to be the judge of what is true or not.[49] At the same time, the social media giant has followed the example of other media platforms in the time leading up to the US election and made a statement in January 2020 saying that they will remove deepfakes and video material that has not been edited in ways that were obvious to the average person.[50] Facebook has also stepped up their overall efforts to tackle disinformation before the 2020 US election, including initiatives to increase transparency by for example showing more information about the owner of Facebook-pages and label content that has been flagged as false by independent fact-checkers. The identification of false material will be carried out by both human fact-checkers and rely on technological solutions.

Facebook has repeatedly requested more government regulation of their activities, which they deem necessary to mitigate the effects of disinformation.[51] This opinion is also expressed in a white paper released in February 2020, intended to address the issues surrounding content regulation from Facebook's viewpoint. Facebook generally advocates for an approach where the government holds internet companies accountable for having certain systems and procedures in place, which the company sees as the best way to ensure an appropriate balancing of safety, freedom of expression, and other values. By requiring systems such as user-friendly channels for reporting content or external oversight of policies or enforcement decisions, and by requiring procedures such as periodic public reporting of enforcement data, regulation could provide governments and individuals the information they need to accurately judge social media companies' efforts[52]. This approach is a continuation of a shift in the company's stance on regulations, which has softened since the tie of the Cambridge Analytica scandal in 2017.[53] The white paper also outlines a number of questions of importance in the content regulation debate, including if illegal content should be defined in regulation and whether social media companies should be required to meet certain performance targets.

[46] (Cuthbertson, 2020)

[47] (Reuters, 2020)

[48] (Deepfake Detection Challenge, 2020)

[49] (The Guardian, 2020)

[50] (Shead, 2020)

[51] (BBC, 2020)

[52] (Facebook, 2020)

[53] (Hutchinson, 2020)

But even though several social media platforms are imposing some kind of content regulation on deepfakes and manipulated material, the interpretation and definitions leads to differences in the practical application. One example is the manipulated video released by Michael Bloomberg's campaign in the primary election 2020, where an outtake from a debate with the Democratic Party candidates was edited to create a comic effect. Twitter stated that the video, had it come out before their new deepfake-policy entered into force in March 2020, would have been removed under the new guidelines. Facebook, on the other hand, stated that the video would most likely not be removed due to their exception for content deemed as satirical expression.

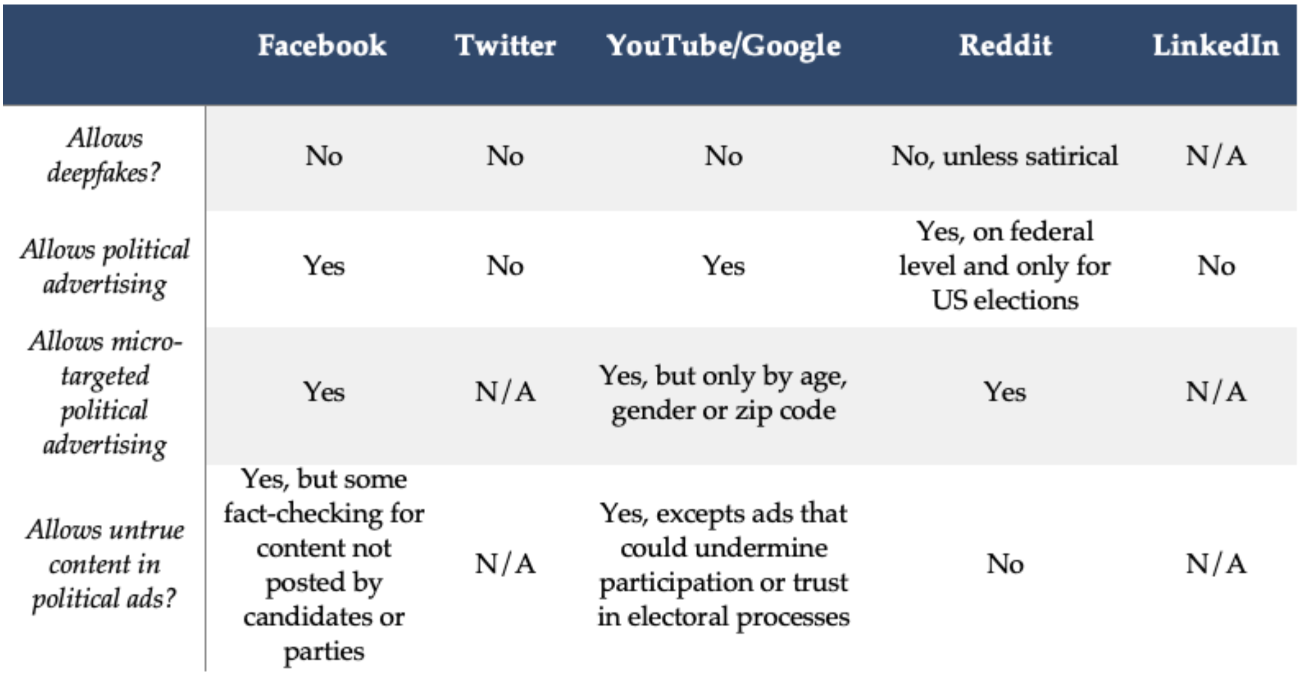
The most important measures taken by a selection of large media platforms have been summarized in the table below (Table 1). The compilation is by no means exhaustive but paints the broad picture of the current situation when it comes to voluntary restrictions by social media platforms.

As seen in the table, large media platforms do already ban parts of the false content in their current shape and form. Still, voluntary commitments made by social media platforms are not legally enforceable by the authorities, and with the previously discussed business model of these platforms, where more traffic means increased revenues, it would be a risk to leave the content moderation completely unregulated in the hands of the multinational corporations. Some states, as will be shown later in this report, have attempted to regulate the social media companies – but there are a number of difficulties in doing so.

Towards regulation – the free speech controversy

Turning our attention to regulatory measures, which would pressure social media companies to take action on disinformation and deepfakes, we encounter several aspects complicating the issue.
A flat out ban on all creation and distribution of deepfakes would, for apparent reasons, be too

Table 1: Voluntary restrictions by platforms



	Facebook	Twitter	YouTube/Google	Reddit	LinkedIn
Allows deepfakes?	No	No	No	No, unless satirical	N/A
Allows political advertising	Yes	No	Yes	Yes, on federal level and only for US elections	No
Allows micro-targeted political advertising	Yes	N/A	Yes, but only by age, gender or zip code	Yes	N/A
Allows untrue content in political ads?	Yes, but some fact-checking for content not posted by candidates or parties	N/A	Yes, excepts ads that could undermine participation or trust in electoral processes	No	N/A

broad. As mentioned in a previous section, deepfake technology can indeed be useful and positive with contribution to for example the clarity of digital content and the film industry.[54] Another struggle is to strike the right balance between content moderation and free speech. The free expression of opinions is a cornerstone of our democracies, and when it comes to disinformation and deepfakes the line between free speech and illegal content is not always clear, especially as deepfakes can be used for humoristic, creative and satirical purposes. How do you prove the intent of a deepfake creator? As stated by Citron and Chesney (2019), to create a law that is prohibiting the harmful deepfakes while still giving room for the expression of free speech and satire is indeed *difficult, but not impossible.*

As much of the literature on election interference stems from the United States, and since many of the social media platforms are based in the US, it's interesting to briefly discuss some of the differences in how free speech is interpreted in Europe compared to the US. While the right of citizens to freely express opinions is a pillar of democracies on both sides of the Atlantic, the interpretations differ. The First Amendment is the leading constitutional guidance in citizen's right to express themselves in the United States, and with its wide scope it for example protects instances of hate speech to some extent – excluding the instances where there is incitement for violence or true threat. European states, on the other hand, are more restrictive in their freedom to express opinions as European case law has deemed some restrictions necessary to protect "the right of others".[55] The discussion is not only of academic relevance, as research has shown significant differences in the attitudes of citizens in Europe and the United States when it comes to offensive statements, where US citizens are generally more tolerant in their interpretation of free speech than for example Germans. But this difference is also reflected in constitutional rights, as there is a much greater protection for publishers, and thereby also social media platforms, in the US constitution than in the European counterparts. While European states in some instances have the possibility to impose content regulation on companies, American legislators are leaning more towards greater transparency and data protection from the tech sector to not be unconstitutional. Potential transatlantic cooperation on legislation would therefore need to fall in the latter category.

The EU response to disinformation

Since at least the 2016 US presidential election, legislative and regulatory measures have been considered at various political levels in order to curb the possible dangers of disinformation on social discourse – ranging from voluntary commitments from social networks, search engines and other companies to guidelines for moderating and deleting illegal content and new supervisory authorities.[56] There are advantages and disadvantages associated with all approaches, and while some states have been more aggressive in their attempts to regulate these issues by introducing legislation, the perhaps broadest measures to date are the ones taken by the social media platforms themselves.

Despite the fact that disinformation is no new issue, governments, international organizations

[54] (Chesney & Citron, 2019)

[55] (Jaursch, Lenoir, Schafer, & Soula, 2019)

[56] (Jaursch, 2019)

and the European Union have only recently started to develop more direct strategies to target these problems. This section will look into Europe's response to the increased occurrence of disinformation campaigns by examining the current initiatives taken by both the EU and a number of key member states within the union on the issue, namely Germany, France, and the United Kingdom. The efforts can largely be divided into three categories; *(I) Regulatory measures*, where legislation binds media platforms to remove illegal content or meet other specified criteria, *(II) Co-regulatory measures*, where media platforms to different extents are recommended to voluntarily bind themselves to fulfill certain requirements and *(III) Public awareness or educational measures*, where the states are informing actors on the topic to proactively mitigate the effects of disinformation. A predominant focus will be put regulatory measures, but as the legislative approach is far from always prioritized, the section will also highlight other initiatives taken by the EU the member states.

The European Union

The first initiatives to tackle disinformation from the EU came in 2015 with the creation of the *East StratCom Task Force*, a part of the European External Action Service. The team is tasked with developing communication material and campaigns to raise awareness and update the public on disinformation in society, primarily stemming from Russia. While the task force was a first step in dealing with the increased threat and widely welcomed by member states, the issue of disinformation was still not high on the political agenda. In a 2018 report, interviews with EU officials are quoted saying that the political support within the EEAS and from the High Representative to deal with information was weak.[57] The change after the US election in 2016 and the Brexit referendum is noticeable, serving as a wake-up call for the EU institutions. With concerns about foreign election interference in the 2019 European election, the EU decided to accelerate the pace of its efforts to counter external disinformation threats.

The central piece is the *Action Plan against Disinformation* launched in December 2018 by the European Commission, answering to the European Council's call for measures to "protect the Union's democratic systems and combat disinformation, including in the context of the upcoming European elections".[58] Although the Action Plan specifically mentions a focus on the 2019 European Election, the document proposes multidimensional and long-term strategies. Disinformation is defined as *verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public*, but excludes satire and parody or clearly partisan news and commentary.[59] Generally, the EU takes more of a defensive than offensive approach in tackling disinformation, as efforts are mostly *responding* to disinformation campaigns rather than *preventing* them.[60] The actions to counter disinformation are based on four pillars and listed in Table 2.

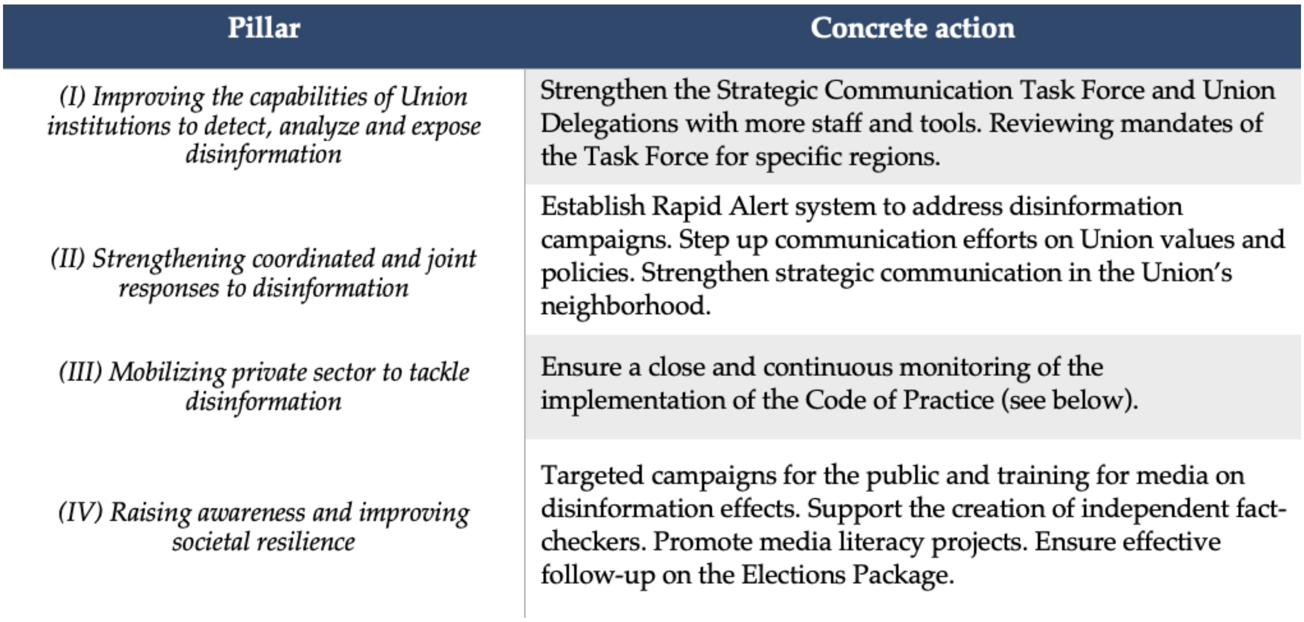
[57] (Scheidt, 2019)

[58] (European Commission, 2018)

[59] (Ibid.)

[60] (Scheidt, 2019)

Table 2: The four pillars of the Action Plan against Disinformation



Pillar	Concrete action
(I) Improving the capabilities of Union institutions to detect, analyze and expose disinformation	Strengthen the Strategic Communication Task Force and Union Delegations with more staff and tools. Reviewing mandates of the Task Force for specific regions.
(II) Strengthening coordinated and joint responses to disinformation	Establish Rapid Alert system to address disinformation campaigns. Step up communication efforts on Union values and policies. Strengthen strategic communication in the Union's neighborhood.
(III) Mobilizing private sector to tackle disinformation	Ensure a close and continuous monitoring of the implementation of the Code of Practice (see below).
(IV) Raising awareness and improving societal resilience	Targeted campaigns for the public and training for media on disinformation effects. Support the creation of independent fact-checkers. Promote media literacy projects. Ensure effective follow-up on the Elections Package.

The European Commission also presented their Code of Practice on Disinformation in April 2018, aimed to address the spread of online disinformation and fake news by voluntary self-regulation by online platforms, social networks and advertisers.[61] As the first worldwide industry agreement on the issue of disinformation, the code sets out a wide range of commitments, from transparency in political advertising to the closure of fake accounts, that the platforms could agree to follow. At this point in time, large social media corporations like Facebook, Google, Twitter and Microsoft have all signed the Code of Practice. In October 2019, one year after its initiation, the signing companies were bound to submit self-assessment reports of the implementation and progress. The European Commission concluded that the companies had increased transparency and had a better dialogue with the EU regarding their policies against disinformation. The scope of action undertaken by each platform varies significantly, however, and there are differences in the implementation of the code between member states. The Commission had specifically monitored the companies during the period leading up to the European election and a clear progress in the commitments could be noted. At the same time, the commitments to empower consumers and provide data for the research community still lacked – leading to a somewhat mixed assessment of the code after its first year. The Commission is releasing a comprehensive assessment of the initiative in the first half of 2020.[62]

There are also discussions about introducing legislation on the issue, mainly by the planned Digital Services Act which is currently being drafted. The goal of the Act would be to update the "liability and safety rules for digital platforms, services and products". Early concept papers have confirmed that the liability rules would be central in the new legislation, but the precise content of this new act is still unspecified, making it difficult to examine at this stage.[63] Currently, the *e-Commerce directive* 2000/31/EC provides the legal framework

[61] (European Commission, 2018)

[62] (European Commission, 2018)

[63] (Propp, 2019)

for online services in the internal market, but as it was drafted as early as 2000, before the digital landscape of today emerged, and exempts digital platforms from content liability after the content has been published (provided illegal content is removed after being brought to their attention), the need for reform is apparent.[64]

Short comment: The EU efforts mainly rely on voluntary, self-regulatory measures and information, which have been criticized with reference to their lack of enforcement. There is a discrepancy between the social media platforms actions and the EU expectation. Deepfakes are not targeted specifically in the EU policy as it focuses on disinformation in a broader context but would fall under the definition of disinformation as it is written in the Action Plan. The Digital Services Act could be set to address the legislative space on the issue of disinformation, but the specific content remains unknown to date.

Germany

The most relevant parts of the German legislation to tackle disinformation and deepfakes are based on rules for removal of illegal content on the internet. As previously mentioned, the e-Commerce Directive covers these issues at the EU level, whereas in Germany, it is the Network Enforcement Act (NetzDG) that regulates illegal content on social media platforms. Informally known as the "hate speech law" or the "Facebook law", it is perhaps the most ambitious attempt by a Western state to hold social media platforms responsible for illegal content,[65] also serving as an inspiration for the French legislation on these issues. The federal law, adopted by the Bundestag in June 2017, aims to improve law enforcement and increase responsibility for social media platforms to act on online speech that is punishable under domestic criminal law. As stipulated in the NetzDG, the platforms must provide users with a mechanism to report illegal content, investigate it, and delete material within 24 hours after notification if it proves to be "manifestly unlawful".[66] Platforms risk a fine of up to €50 million if they don't comply with these rules, but are not obliged to actively search for illegal content themselves.[67] The NetzDG also imposes *transparency* requirements for platforms, meaning that the companies have to publish reports on their content moderation practices if they receive more than a pre-determined number of complaints per year for users.

The NetzDG is an example of a state focusing on the content moderation approach, and its clear and transparent rules for social media platforms is definitely a helpful tool compliant with current legislation. But while the NetzDG largely fills its general purpose of providing the legal base for removal of illegal content, it also raises the question about *what* content should be classified as illegal and where the line between free speech and criminal conduct should be drawn. Whether intended or not, the determination of law breaching is now outsourced to the media platforms. The law has become very divisive and heavily criticized by both the affected social media platforms and some scholars for threatening free speech in Germany and being unconstitutional in its design.[68] While it is still early to assess the full impact of the NetzDG, initial reports seems to suggest that the risk of over-blocking content has been overestimated, but also that the impact of the

[64] (Quain & Cheallacháin, 2019)

[65] (Tworek, 2019)

[66] (Ibid.)

[67] (Jaursch, 2019)

[68] (Schmitz-Berndt, M. Berndt, Rechtsanwalt, & Rechtsanwälte)

regulation when it comes to removal of content has not been substantially different from the normal community standards.[69]

In addition to the NetzDG, there are at least two other regulatory areas that may be relevant to curb disinformation and deepfakes in Germany, even though it is not the main purpose of these regulations. The planned *Interstate Media Treaty*, an addition to the current Interstate Broadcasting Treaty regulating the traditional TV- and radio broadcasting in Germany, is currently discussed to extend its focus to online media. The treaty is intended to create an information environment that does not hinder democratic discourse, but with uncertainties about the final design and scope of the amendment, it may be too optimistic to expect anything more than enforced rules on transparency for online media companies as the best possible outcome.[70] The other relevant area in German legislation regards *competition and privacy policy*. This option would not tackle disinformation in itself, but rather limit the ability for media platforms to collect data from their users to create segmented news feeds. Enforcement of this option is difficult, however, as data protection authorities often lack the necessary resources. It should also be mentioned that, apart from the NetzDG, many of the German laws and reforms that deals with society and economy in the digital age are not designed to handle disinformation as we know it from 2016 and onwards. Reforms on the digitalization of economy in competition law, the use of data protection with regards to personal data and broadcasting regulations have been discussed for a long time, and it is unclear how they can be adapted to the new disinformation environment.[71]

Short comment: With its central legislation to tackle disinformation (and potentially also deepfakes specifically) being the NetzDG, Germany does have some regulatory capacity to limit the spread of illegal content. The questions of free speech and liability for platforms remain, however. The regulatory framework is also fragmented, and a holistic approach to tackle disinformation and deepfakes is lacking. In addition, the NetzDG does not regulate false information or fake news per sé, but rather what is deemed to be illegal content.

France

There are two main initiatives for content regulation in France. The most apparent one relating to disinformation is the law against manipulation of information, commonly referred to as the "Fake News bill". Intended to better protect democracy towards the different ways in which fake news is deliberately spread,[72] it was approved in November 2018 and has caused quite some controversy, mainly around concerns of press censoring.[73] It is considered western Europe's first attempt to officially ban false information online and is in accordance with French President Macron's beliefs that internet must be regulated. The law is specifically targeting election campaigns and can therefore only be applied three months before an election. During this period, any individual, political party or public authority can appeal to a judge for false content it considers should be removed from a social media platform. The judge then has 48 hours to act upon the appeal.[74] In addition, platforms such as Facebook, Twitter and YouTube are required to increase *transparency* around political advertising.

[69] (Ibid.)

[70] (Jaursch, 2019)

[71] (Ibid.)

[72] (French Government, 2018)

[73] (Fiorentino, 2018)

[74] (Funke & Flamini, 2019)

This means that the companies must publish who has purchased sponsored content or campaign ads and the price for which it has been sold.[75]

Part of the legislation issues the French Higher Audiovisual Council (CSA) to publish a report following up on how well the companies are abiding the law and gives them the ability to revoke broadcaster rights of TV and radio outlets in certain cases.[76] The media platforms must also publicly introduce measures to combat fake news and allow their users to bring attention to information they believe to be fake. The bill defines content labelled as fake news, as it must be *(I) manifest*, *(II) disseminated deliberately* on a large scale and *(III) lead to a disturbance of peace or compromise the results of elections*. Fines and imprisonment are the consequences facing the social media platforms if the obligations are not met.[77]

In addition to this, a new suggested bill on hate speech has drawn a lot of media attention in the last year. The bill proposes stricter demands on social media platforms to screen content that is discriminatory and gives a 24-hour deadline for platforms to remove reported content, much like the German NetzDG. Big tech companies would also have to hire extra moderators to properly screen the content. Large fines will be imposed on the tech platforms if the requirements for screening is not met, and while the law is neither targeting deepfakes specifically, nor has it been implemented, it is another sign of France's aggressive regulatory approach to tackle disinformation. Just before a meeting between French President Macron and Facebook CEO Mark Zuckerberg in May 2019, a French report calling for more access to Facebook algorithms was released with President Macron stating that France should take a leading role in tech regulation.[78] The report came after Facebook allowed a team of French regulators to spend half a year monitoring the companies policies from the inside.

Short comment: France is taking quite aggressive action to impose regulatory measures to prevent spreading of disinformation. While the bills, just as in Germany, have received criticism with regards to the free speech versus content moderation debate, France is undoubtedly trying to take leadership in regulating social media platform and their actions surrounding elections. While deepfakes are not specifically mentioned in the Fake news-bill, it is a more direct way of targeting misinformation than the legislation on hate speech.

The United Kingdom

While there is no explicit legislation prohibiting online publication of disinformation in the United Kingdom, the government is currently investigating both the impact of the issue and the possibilities for regulation. Having experienced foreign disinformation campaigns both surrounding the Brexit referendums and in the aftermaths of the poisoning of Sergei Skripal, the British Prime Minister announced that the intelligence services would be responsible for identifying social media platforms distribution disinformation under the newly introduced Fusion Doctrine.[79] The Fusion Doctrine is intended to improve the ability of the National Security

[75] (Funke & Flamini, 2019)

[76] (Ibid.)

[77] (Jaursch, Lenoir, Schafer, & Soula, 2019)

[78] (France 24, 2019)

[79] (Feikert-Ahalt, 2019)

Council to make national security strategy and implement the decisions across government.[80] Several departments within the government were also tasked with investigating the impact of disinformation and provide policy recommendations on how to tackle the issues.

The result of the 18-month investigation into disinformation and fake news was a final report released in May 2019. The report heavily criticized Facebook for failing to address the attempts by Russia to manipulate elections, going as far as to call the social media platform "digital gangsters".[81] Part of the critique stems from accusations that Facebook choses profit over user's data security and privacy rights. The reports also came with a number of recommendations, including propositions of comprehensive new regulations tightening tech companies' liabilities for published content and the removal of illegal material. The government published a response agreeing with the report's conclusion that the current self-regulatory approach is insufficient in tackling online disinformation and that regulations are needed both on that area and in ensuring platforms removal of illegal content,[82] but it has not resulted in any legislation to date. Another report on the Russian interference in the United Kingdom authored by the Parliament's Intelligence and Security Committee was set to be released before the election in December 2019 but was stopped by UK Prime Minister Boris Johnson. The report, controversially delayed by Prime Minister Boris Johnson with reference to the inability to clear the report of sensitive material in time for the election, now seems to be held until a new Intelligence and Security Committee has been appointed, which would mean no earlier than the first half of 2020.[83] The official line has however changed, and UK ministers are now no longer allowed to say that there have been "no successful examples" of Russian interventions.

The United Kingdom does not have a regulatory body overseeing the social media platform and online content as a whole. Neither is there, as stated before, laws prohibiting the spread of disinformation and illicit deepfakes. The closest regulatory unit is the Office of Communications (Ofcom), established under the Communications Act in 2003 to enforce content standards across radio and television. Ofcom has itself argued that there is a lack of legislation to cover online contents, making television and radio publishers follow tough rules that social media platforms can avoid.[84] Under PM Theresa May, the UK government announced their establishment of a task force, or a "dedicated national security communications unit", specifically targeting the disinformation and fake news[85] and in July 2019 the UK government also announced an 18 million pond package to counter disinformation over 3 years,[86] together with plans to teach children in schools about how to spot disinformation online.[87]

Short comment: Despite the reported interference in the Brexit referendum, the United Kingdom has reacted quite slowly to the threat of disinformation, especially considering the new election held in 2019 and the risks involved. The issued reports and response from the government does however suggest that legislation is to be expected.

[80] (UK Parliament, 2019)

[81] (Pegg, 2019)

[82] (Parliament, 2019)

[83] (Sabbagh, 2020)

[84] (Feikert-Ahalt, 2019)

[85] (BBC, 2018)

[86] (UK Government, 2019)

[87] (BBC, 2019)

Recommendations and lessons learned from the EU

Disinformation is no longer a problem hiding in the shadows. Although the majority of all states, including the European Union, reacted slowly to the increased threat of digital foreign election interference, there has been a clear and well-needed intensification of efforts to tackle these issues in recent years. The voluntary restrictions taken by social media platforms can go a long way in mitigating the effects of disinformation campaigns and it is therefore crucial to work *with* the platforms, and not *against* them.

Recommendation 1: Intensify collaboration between states and platforms

- Collaboration between states and the media platforms on these issues, keeping an open dialogue on solutions and ways to improve and design measures, will most likely lead to a better result than an approach where the social media platforms are depicted as the main villains, which runs the risk of alienating the perhaps most important actor in the fight against disinformation.
- Voluntary restrictions should be encouraged, as they are an important tool in mitigating the spread of disinformation.
- While voluntary restrictions are important, they should by no means be the *only* measure restricting the spread of false content online. States could still have a demanding stance on the platforms' policies regarding content moderation, especially considering the business models of these companies and the algorithms they are built upon.

There is also a broader legal and philosophical question to be posed about the role of large, multinational corporations in today's society, increasingly taking on responsibilities and rights previously held exclusively by states. The less hard regulation the media platforms are faced with, the larger part they will take in judging what is legal an illegal in the societal and political debate of today – a type of privatized enforcement of the law. As even Facebook themselves are now requesting harder regulation to not have these considerations outsourced to multinational corporations, it is safe to say that regulatory efforts from states needs to be stepped up even more. The free speech balancing act does however still pose a real issue for regulators, and the concern of over-removal of content is in itself a risk for our democratic societies. The French Fake news bill, specifically targeting the period surrounding an election, is perhaps the most promising measure at this point, as it is event-limited and focuses on *transparency*.

Recommendation 2: Transparency is key in regulating digital disinformation

- Increased transparency is a policy-objective that applies to all aspects of the issue, from political advertising to the performance requirements of the platforms, and should therefore always be part of legislation.
- Transparency rules can be used on both sides of the Atlantic, regardless of free speech interpretation and tradition. The use of a *neutral, state-appointed judge* to determine the intent and illegality of content is another key, and inspiration can be taken from France.
- The party or organization behind political ads should always be clear. Social media platforms should consider only allowing actors who have pledged to avoid deceptive campaign practices

to purchase ads. Such pledges, like *The Pledge for Election Integrity* by the Transatlantic Commission on Election Integrity, should then become working standards for platforms to decide on whether to accept any given ad.

- Platforms should also provide data on their practices and removal of illegal content.

Deepfake technology is indeed a looming challenge for democracies and may play a part already in the US 2020 election. Direct efforts to tackle deepfakes have so far mainly been taken by platforms themselves, and initiatives like the Deepfake Detection challenge shows that the companies are continuing to develop their deepfake strategies. Going forward, it is crucial that legislation targeting disinformation is formulated to also capture deepfakes.

Recommendation 3: Removal of manipulated material must be coherent and well-defined

- Manipulated material and deepfakes must be addressed to avoid its potential implications on democratic elections. Definitions of *illegal* should focus on illicit intent and content that has been manipulated to an extent where it is not obvious to the viewer.
- Acknowledging that the above definition opens for subjectivity and given that the platforms need to urgently make decisions on whether to remove content believed to be manipulated, platforms could *always flag content as manipulated as a first step*, to give more time for an assessment of intention.
- Labelling material as manipulated is easier than proving intention, and while content will not be immediately removed, a clear labelling of material and manipulated will raise awareness among viewers until a final decision of removal has been made.

EU's softer approach, avoiding hard regulation, has been criticised for its lack of enforcement, but the complexity introducing legislation on an EU level, widely exceeding that of legislation in individual member states, should be kept in mind. The fact that EU has taken a holistic approach and upped their efforts sends a clear signal to the member states that the issue is prioritized. An update of the e-Commerce directives should nonetheless be a top-priority, and the indications that the new Digital Services Act will increase liability of social media platforms is without doubt a welcomed update to bring the EU's regulatory measures in the digital sphere into the 2020's. At the same time, the informative approach should not be underestimated.

Recommendation 4: Educational material should accompany legislation and voluntary restrictions

- Educating voters, citizens and users of social media platforms is an easy, relatively cheap and effective way to increase awareness about disinformation.
- Specific funds should be allocated by states, the EU and social media companies to enable increased awareness campaigns and to support organizations, initiatives and projects focused on media literacy.

Summing up, the potential effect disinformation may have on our democratic societies in the digital age is unquestionable. While efforts to tackle the issue have been stepped up by both private actors and states, so has the sophistication of coordinated attempts to interfere in elections. As there is no indication that the trend is turning, efforts to tackle disinformation now also needs to enter the new decade to protect our democracies.

@AoDemocracies

Bibliography

Alba, D., & Satariano, A. (2019, September 26). At Least 70 Countries Have Had Disinformation Campaigns, Study Finds. *New York Times*.

Aspinwall, N. (2020, January 10). Taiwan's War on Fake News Is Hitting the Wrong Targets. *Foreign Policy*.

Bayer, J., Bitiukova, N., Bárd, P., Szakács, J., Alemanno, A., & Uszkiewicz, E. (2019). *Disinformation and propaganda – Impact on the functioning of the rule of law in the EU and its Member States.* European Parliament.

BBC. (2018, January 23). *Government announces anti-fake news unit*. Retrieved from BBC News: https://www.bbc.com/news/uk-politics-42791218

BBC. (2019, July 15). *Fake news and how to spot it to be taught in schools*. Retrieved from BBC Newsround: https://www.bbc.co.uk/newsround/48988778

BBC. (2019, October 31). *Twitter to ban all political advertising*. Retrieved from BBC News: https://www.bbc.com/news/world-us-canada-50243306

BBC. (2020, February 15). *Mark Zuckerberg: Facebook boss urges tighter regulation*. Retrieved from BBC News Online: https://www.bbc.com/news/technology-51518773

Chandler, S. (2020, March 9). *Why Deepfakes Are A Net Positive For Humanity*. Retrieved from Forbes: https://www.forbes.com/sites/simonchandler/2020/03/09/why-deepfakes-are-a-net-positive-for-humanity/#9d2b3872f84f

Chertoff, M., & Donahoe, E. (2018, November 12). *For election hackers, a new and more dangerous tool.* Retrieved from Reuters: https://www.reuters.com/article/chertoffdonahoe-deepfakes/column-for-election-hackers-a-new-and-more-dangerous-tool-idUSL2N1XN14R

Chesney, B., & Citron, D. (2019). *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security.* California Law Review.

Corstange, D., & Marinov, N. (2012). Taking Sides in Other People's Elections: The Polarizing Effect of Foreign Intervention. *American Journal of Political Science*(Volume 56, Issue 3), pp. 655-670.

Cuthbertson, A. (2020, February 5). *Twtter bans deepfakes and deceptive media ahead of US elections.* Retrieved from Independent: https://www.independent.co.uk/life-style/gadgets-and-tech/news/twitter-ban-deepfake-us-election-policy-trump-2020-a9318276.html

Deepfake Detection Challenge. (2020). *Deepfake Detection Challenge*. Retrieved from deepfakedetectionchallenge.ai: https://deepfakedetectionchallenge.ai

Edelman, G. (2020, Janaury 7). *Facebook's Deepfake Ban Is a Solution to a Distant Problem*. Retrieved from Wired: https://www.wired.com/story/facebook-deepfake-ban-disinformation/

European Commission. (2018, December 5). *Action Plan against Disinformation*. Retrieved from HIGH REPRESENTATIVE OF THE UNION FOR FOREIGN AFFAIRS AND SECURITY POLICY: https://eeas.europa.eu/sites/eeas/files/action_plan_against_disinformation.pdf

European Commission. (2018, September). *Code of Practice on Disinformation.* Retrieved from https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation

European Commission. (2018, September). *Code of Practice on Disinformation.* Retrieved from https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation

European Parliament. (2019). *Online disinformation and the EU's response.* Retrieved from http://www.europarl.europa.eu/RegData/etudes/ATAG/2018/620230/EPRS_ATA(2018)620230_EN.pdf

European Parliament. (2020). *European Parliament portal.* Retrieved from http://www.europarl.europa.eu/RegData/etudes/ATAG/2018/620230/EPRS_ATA(2018)620230_EN.pdf

Facebook. (2020, February). *Charting A Way Forward: Online Content Regulation.* Retrieved from https://about.fb.com/wp-content/uploads/2020/02/Charting-A-Way-Forward_Online-Content-Regulation-White-Paper-1.pdf

Feikert-Ahalt, C. (2019, April). *Initiatives to Counter Fake News: United Kingdom.* Retrieved from Library of Congress: https://www.loc.gov/law/help/fake-news/uk.php#III

Fiorentino, M.-R. (2018, November 22). *France passes controversial 'fake news' law*. Retrieved from Euronews: https://www.euronews.com/2018/11/22/france-passes-controversial-fake-news-law

France 24. (2019, May 10). *French report calls for more access to Facebook algorithms as Macron meets Zuckerberg*. Retrieved from France 24 : https://www.france24.com/en/20190510-france-facebook-law-mark-zuckerberg-president-macron-internet-regulation-internet

Freedom House. (2017). *Manipulating Social Media to Undermine Democracy*. Retrieved from Freedom on the Net 2017: https://freedomhouse.org/report/freedom-net/freedom-net-2017

French Government. (2018). *Against information manipulation*. Retrieved from Gouvernement.fr: https://www.gouvernement.fr/en/against-information-manipulation

Funke, D., & Flamini, D. (2019). *A guide to anti-misinformation actions around the world.* Retrieved from Poynter.org: https://www.poynter.org/ifcn/anti-misinformation-actions/#sweden

Godinez, J. (2018). The Vested Interest Theory: Novel Methodology Examining US-Foreign Electoral Intervention. *Journal of Strategic Security*(11 (2)), pp. 1-31.

Hutchinson, A. (2020, February 18). *Facebook Publishes New Whitepaper on Standardized Online Content Regulation*. Retrieved from Social Media tToday: https://www.socialmediatoday.com/news/facebook-publishes-new-whitepaper-on-standardized-online-content-regulation/572416/

Jaursch, J. (2019). *Regulatory Reactions to Disinformation - How Germany and the EU are trying to tackle opinion manipulation on digital platforms.* Stiftung Neue Verantwortung.

Jaursch, J. (2019). *Regulatory reactions to disinformation.* Stiftung Neue Verantwortung.

Jaursch, J., Lenoir, T., Schafer, B., & Soula, E. (2019, November 15). *Tackling Disinformation : Going Beyond Content Moderation*. Retrieved from Institut Montaigne: https://www.institutmontaigne.org/en/blog/tackling-disinformation-going-beyond-content-moderation

Kofi Annan Foundation. (2020). *Protecting Electoral Integrity in the Digital Age.* Retrieved from https://www.kofiannanfoundation.org/app/uploads/2020/01/f035dd8e-kaf_kacedda_report_2019_web.pdf

Landau, S. (2018). *Cybersecurity: Time for a New Definition*. Retrieved from Lawfare: https://www.lawfareblog.com/cybersecurity-time-new-definition

Levin, D. H. (2016, June). When the Great Power Gets a Vote: The Effects of Great Power Electoral Interventions on Election Results. *International Studies Quarterly*(Volume 60), pp. 189–202.

Levin, D. H. (2018). A Vote for Freedom? The Effects of Partisan Electoral Interventions on Regime Type. *Journal of Conflict Resolution*(63 (4)), pp. 839–868.

Lührmann, A., Grahn, S., Pillai, S., & Lindberg, S. I. (2019). *Democracy Facing Global Challenges V-DEM ANNUAL DEMOCRACY REPORT 2019.* Gothenburgh: University of Gothenburg.

Metzger, M. (2019, November 15). *Effectiveness of Responses to Synthetic and Manipulated Media on Social Media Platforms*. Retrieved from Carnegie Endowment for International Peace: https://carnegieendowment.org/2019/11/15/legal-ethical-and-efficacy-dimensions-of-managing-synthetic-and-manipulated-media-pub-80439#effectiveness

Murphy, C., & Huang, Z. (2019, September 2). *China's Red-Hot Face-Swapping App Provokes Privacy Concern.* Retrieved from Bloomberg: https://www.bloomberg.com/news/articles/2019-09-02/china-s-red-hot-face-swapping-app-provokes-privacy-concern

Nilesh, C. (2020, February 18). *We've Just Seen the First Use of Deepfakes in an Indian Election Campaign*. Retrieved from Vice News: https://www.vice.com/en_in/article/jgedjb/the-first-use-of-deepfakes-in-indian-election-by-bjp

Palmer, A. (2019, October 1). *Facebook, Twitter accuse China of running disinformation campaign against Hong Kong protesters*. Retrieved from CNBC: https://www.cnbc.com/2019/08/19/twitter-accuses-china-of-running-disinformation-campaign-against-hong-kong-protesters.html

Parliament, U. (2019, May 9). *Disinformation and 'fake news': Final Report: Government Response to the Committee's Eighth Report*. Retrieved from UK Parliament publication: https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/2184/218402.htm

Pegg, D. (2019, February 18). *Facebook labelled 'digital gangsters' by report on fake news*. Retrieved from The Guardian: https://www.theguardian.com/technology/2019/feb/18/facebook-fake-news-investigation-report-regulation-privacy-law-dcms

Propp, K. (2019, October 22). *The emerging EU regulatory landscape for digital platform liability.* Retrieved from The Atlantic Council: https://www.atlanticcouncil.org/blogs/new-atlanticist/the-emerging-eu-regulatory-landscape-for-digital-platform-liability/

Quain, C., & Ceallacháin, D. N. (2019, November 28). *DISINFORMATION IN THE DIGITAL AGE: THE NEED TO REGULATE*. Retrieved from The Institute for International and European Affairs: https://www.iiea.com/digital/digital-blog/disinformation-in-the-digital-age/

Reuters. (2020, February 3). *YouTube reinforces guidelines on fighting misleading election content*. Retrieved from Reuters : https://www.reuters.com/article/alphabet-election/youtube-reinforces-guidelines-on-fighting-misleading-election-content-idUSL4N2A3387

Sabbagh, D. (2020, January 20). *Boris Johnson urged to publish report on Russian meddling*. Retrieved from The Guardian: https://www.theguardian.com/politics/2020/jan/20/boris-johnson-urged-to-publish-report-on-russian-meddling

Sabbagh, D. (2020, March 15). *UK ministers will no longer claim no successful examples of Russian interventions*. Retrieved from The Guardian: https://www.theguardian.com/technology/2020/mar/15/uk-ministers-will-no-longer-claim-no-successful-examples-of-russian-interference

Scheidt, M. (2019). *The European Union versus External Disinformation Campaigns in the Midst of Information Warfare: Ready for the Battle?* EU Diplomacy Paper, College of Europe, DEPARTMENT OF EU INTERNATIONAL RELATIONS AND DIPLOMACY STUDIES, Brugge.

Schmitz-Berndt, S., M. Berndt, C., Rechtsanwalt, D. S., & Rechtsanwälte, B. (n.d.). *The German Act on Improving Law Enforcement on Social Net- works: A Blunt Sword?* Universite ́du Luxembourg.

Schneier, B. (2020, January 7). *Bots Are Destroying Political Discourse As We Know It*. Retrieved from The Atlantic: https://www.theatlantic.com/technology/archive/2020/01/future-politics-bots-drowning-out-humans/604489/

Scott, M. (2019, May 11). UK ignores warnings of digital election interference. *Politico.*

Shead, S. (2020, January 7). *Facebook to ban 'deepfakes'*. Retrieved from BBC News: https://www.bbc.com/news/technology-51018758

The Guardian. (2020, January 9). *Facebook refuses to restrict untruthful political ads and micro-targeting.* Retrieved from The Guardian Online: https://www.theguardian.com/technology/2020/jan/09/facebook-political-ads-micro-targeting-us-election

Tworek, H. (2019). *An Analysis of Germany's NetzDG Law.* Transatlantic Working Group.

UK Government. (2019, July 7). *UK steps up fight against fake news*. Retrieved from Gov.uk: https://www.gov.uk/government/news/uk-steps-up-fight-against-fake-news

UK Parliament. (2019, July 21). *Optimising the national security strategy-making process*. Retrieved from UK Parliament Publications: https://publications.parliament.uk/pa/jt201719/jtselect/jtnatsec/2072/207206.htm

Vosoughi, S. e. (2018). The Spread of True and False News Online. *359 SCIENCE 1146*.