INFORMING THE DISINFO DEBATE:

A POLICY GUIDE FOR PROTECTING HUMAN RIGHTS

December 2021

accessnow

CIVIL LIBERTIES UNION FOR EUROPE

EDRi

Authors

Eliška Pírková, Access Now

Filip Lukáš, Access Now

Eva Simon, Civil Liberties Union for Europe

Franziska Otto, Civil Liberties Union for Europe

Diego Naranjo, EDRi

The co-authors are grateful for the insights from the following organisations to the report: Asociația pentru Tehnologie și Internet (Apti), Bits of Freedom, Panoptykon, Citizen D, Naphsica Papanicolaou (Wikimedia France), IT-Pol, and XNet

Table of contents

I. Executive Summary

The 2016 US presidential election and the Brexit vote campaigns shed light on how impactful online disinformation and propaganda can be.[1] EU policymakers sought solutions to mitigate the effects of online disinformation ahead of the 2019 EU elections, continuing to do so during the COVID-19 pandemic and in preparation for the 2024 European elections. However, it is not the phenomenon of disinformation that is a novelty, but the role digital technologies play in helping to create, disseminate, and amplify disinformation.

The complexity of tackling disinformation has been widely recognised by academia, policy makers, civil society organisations, and human rights advocates. Disinformation and online disinformation is a multifaceted societal issue that cannot be resolved with quick fixes. So far the European Union has failed to put forward effective solutions. The EU's focus — including in the Code of Practice on Disinformation[2] — on identifying concrete categories of online content for removal and promoting metrics of success that include the quick takedown of a high quantity of content has clearly missed the mark. The approach is evolving as there are now a number of legislative proposals in the EU and beyond for a regulatory response that targets how content is being distributed, personalised, and curated by very large online platforms as part of manipulative, data-driven business strategies to increase profit. These proposals include the recently launched, EU-proposed Regulation on political advertisement.[3]

This joint report is the continuation of its 2018 predecessor, *Informing the "Disinformation" Debate*.[4] The 2018 report is among the first by civil society organisations to point to platforms' problematic business models as a fundamental factor behind the online manipulation of people's economic and political choices. There is now a growing consensus that regulatory approaches must address the business model as a foundational matter, as a large number of policy analyses argue.[5] In this report, we unpack the main methods of manipulation

1 Benkler, Y., Faris, R, Roberts, H. Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics, Oxford University Press, 2018.

2 European Commission, Code of Practice on Disinformation, December 2, 2021.

3 European Commission, European Democracy: *Commission sets out new laws on political advertising, electoral rights and party funding*, November 25, 2021.

4 Access Now, Civil Liberties Union for Europe, and European Digital Rights, *Informing the Disinformation debate*, October 18, 2018.

5 European Commission, *European Democracy Action Plan*, 2021.

that platforms engage in that harm fundamental rights. These methods stem directly from the platforms' business models and have severe impact on the absolute freedom to form an opinion and freedom of thought. They are:

- **Surveillance-based advertisement, including political advertising; and**
- **Amplification of disinformation online via content recommender systems and personalisation of news content.**

The following analysis is informed by our previous works in the field of data protection, privacy, freedom of expression and opinion, and freedom of access to information. **The main outcome of this report is a set of policy recommendations addressed to the EU co-legislators focusing on: how to effectively mitigate fundamental rights risks that result from the manipulative methods deployed by large online platforms that exploit people's vulnerabilities and their sensitive data; and how to combat disinformation in a manner that is fully compliant with fundamental rights standards.**

In order to properly address the problem of disinformation online, we ground our analysis on the following premise: disinformation on online platforms is not the cause but rather a symptom of broader societal problems, such as the dysfunction of politics, power imbalances, inequalities, racism, sexism, and other systems of oppression. Addressing these issues in line with the legally binding EU Charter of Fundamental Rights requires creating the conditions to ensure that large online platforms' business models that fuel disinformation are radically transformed.

It is crucial to fully understand the timeliness and relevance of this joint report. The 2018 report was published in response to several policy documents by different EU bodies and institutions dealing with online disinformation (or "fake news"). Since then, the European Commission has launched several key documents that seek to combat disinformation online, including the European Democracy Action Plan[6] and the revision of the 2018 Code of Practice on Disinformation (preceded by the Guidance on strengthening the Code). All of these efforts are happening against the backdrop of the EU developing the first-of-its-kind horizontal regulation establishing a new model of platform governance, the Digital Services Act (DSA). We urge EU co-legislators to adopt a holistic approach when developing a new model of human rights-centric platform governance that consists of effective enforcement of existing legislation, mainly the General Data Protection Regulation (GDPR); swift adoption of the proposed e-Privacy Regulation; and making sure that fundamental rights safeguards are fully incorporated in negotiations of the DSA and Digital Markets Act (DMA).

6 European Commission, *European Democracy Action Plan: making EU democracies stronger*, December 3, 2020.

II. Introduction: Scope of the Problem

Online platforms as well as state actors have been battling misinformation and disinformation for a long time. A few very large online platforms have gained major influence and ability to shape public discourse. Due to their growing influence, they enabled a new pathway for amplification of disinformation and manipulation of people that has unprecedented reach. In the words of the UN Special Rapporteur on the Protection of the Right to Freedom of Expression and Opinion, this alarming issue is "politically polarising, hinders people from meaningfully exercising their human rights. and destroys their trust in Governments and institutions".[7]

Whistleblower Frances Haugen highlighted in her revelations[8] how Facebook's (now Meta) content ranking via content recommender systems had led to the spread of disinformation and hate speech. In her testimony delivered to subcommittees of the US senate, she explained how her former employer Facebook/Meta was willing to use hateful and harmful content on its site to keep users coming back and to boost users' engagement. Documents from the "Facebook Papers" that she disclosed to the media show the degree to which the company knew of extremist groups on its site trying to polarise US voters before the election.[9] They also reveal that internal researchers had repeatedly determined how its key features amplified toxic content on the platform. The company performs algorithmic curation and personalisation of online content using content recommender systems that leverage machine-learning models to remove or demote content, but these models are only trained for certain types of content. Haugen said Facebook/Meta knows: "Engagement-based ranking is dangerous without integrity and security systems."[10]

The EU and its Member States are legally obliged to respect the Charter of Fundamental Rights in mitigating the problem of disinformation. Freedom of expression and the right to access information lie in the core of democratic discourse. Equally, absolute freedom of thought and freedom of opinion have to be safeguarded against any unjustified interferences. In fact, all fundamental rights and freedoms are impacted by States' efforts to tackle societal phenomena such as disinformation,

7 Irene Khan, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, *Disinformation and freedom of opinion and expression*, point 2, April 13, 2021.

8 The Wall Street Journal, *The Facebook Files: A Wall Street Journal Investigation*, October 1, 2021.

9 The New York Times, *The Facebook Papers: Facebook Wrestles With the Features It Used to Define Social Networking*, October 25, 2021.

10 The Guardian, *Facebook's policing of vitriol is even more lackluster outside the US, critics say,* October 17, 2021.

just as they are by private actors' actions or lack thereof.

The online environment is not the root cause of disinformation, although it can intensify the impact of false information on people and democracies. Tracking and the harvesting of personal data are the core of the business model of most online platforms, which is based on monetising information and content of any kind, including disinformation. The EU must address the business model of online platforms that amplifies the impact of disinformation. While the general discourse about disinformation is focusing on the online ecosystem, evidence shows that disinformation, such as discourses spread around the 2016 US election and Brexit vote, is also carried over mainstream media and other actors operating both online and offline.[11]

This report focuses on the issue of how people should be able to receive and impart information and to form their opinions and thoughts. Freedom of expression is one of the core values of democracies. This freedom is not only about protecting information or ideas that are "favourably received or regarded as inoffensive", but also about protecting those that "offend, shock, or disturb the State or any sector of the population".[12] We support the definitions developed and further elaborated by international human rights monitoring bodies, including the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, that define disinformation as false information that is disseminated intentionally to cause serious social harm. Having said that, it must be noted that there is no universally accepted definition of disinformation at the international level due to high complexity and blurry lines among categories of online content. The lack of agreement on what constitutes disinformation, including the frequent and interchangeable use of the term misinformation, reduces the effectiveness of responses.

11 Knight Foundation, *Seven ways misinformation spread during the 2016 election*, October 4, 2018.

12 European Court of Human Rights, *Case of Handyside v. United Kingdom 5493/72*, December 7, 1976.

III. Human Rights Analysis

What fundamental rights are most impacted by the spread of disinformation?

The responses to the problem of disinformation on behalf of States, regulatory bodies, and Big Tech companies should be in line with the international human rights framework. Disinformation is a complex and challenging issue that has an impact on several fundamental rights. Below, we analyse the ones that are primarily concerned. However, this list should be viewed as open-ended and not exhaustive.

3.1 Right to freedom of expression

> ***Charter of Fundamental Rights of the European Union***
>
> ***Article 11***
> ***Freedom of expression and information***
>
> 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.
>
> 2. The freedom and pluralism of the media shall be respected.

The right to freedom of expression in the European Union protects all forms of expression and their content, regardless of its form, its speaker, or the type of medium used for its distribution.[13] Importantly, freedom of expression equally protects information that offend, shock, or disturb[14] and irrespective of the truth or falsehood of the content.[15] This high threshold of protection is a precondition for a democratic society that stands on functioning rule of law, pluralism of information, diversity of opinions, and non-discrimination principle. The right to freedom of expression is not an absolute right. However, the legitimate grounds for its restrictions are strictly defined in the EU Charter of Fundamental Rights as well as in the European Convention on Human Rights. More precisely, they are permissible under three conditions:

13 Monica Macovei, *Freedom of expression – A guide to the implementation of Article 10 of the European Convention of Human Rights*, 2001.

14 European Court of Human Rights, *Case of Handyside v. United Kingdom 5493/72*, December 7, 1976.

15 European Court of Human Rights, *Salov v. Ukraine, application No. 65518/01, judgment*, September 6, 2005.

1. **They must be prescribed by law;**

2. **They must be issued to pursue a legitimate aim; and**

3. **They must be proportionate and necessary in a democratic society.**

The free flow of information is a critical element of freedom of expression and places a positive obligation on States to proactively put information of public interest in the public domain, and promote plural and diverse sources of information, including media freedom.[16]

3.2 Freedom to hold an opinion

> ***Charter of Fundamental Rights of the European Union***
>
> ***Article 10***
> ***Freedom of thought, conscience, and religion***
>
> Everyone has the right to freedom of thought, conscience, and religion. This right includes freedom to change religion or belief and freedom, either alone or in community with others and in public or in private, to manifest religion or belief, in worship, teaching, practice, and observance.
>
> 2. The right to conscientious objection is recognised, in accordance with the national laws governing the exercise of this right.

In its 2011 General Comment on Article 19 of the International Covenant on Civil and Political Rights (ICCPR) that discusses both freedom of opinion and freedom of expression, the UN Human Rights Committee states that "Freedom of opinion and freedom of expression are indispensable conditions for the full development of the person. They are essential for any society. They constitute the foundation stone for every free and democratic society and the freedoms of opinion and expression form a basis for the full enjoyment of a wide range of other human rights".[17]

The international human rights framework distinguishes between the internal and external dimension of the right to freedom of thought, and the closely related right to freedom of opinion. While the external dimension of these fundamental freedoms can be subject to legitimate restrictions that must be necessary in a democratic society, proportionate, and non-discriminatory, the internal dimension of the freedom of thought and freedom of

16 Irene Khan, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, *Disinformation and freedom of opinion and expression*, 2021.

17 UN Human Rights Committee, General Comment No. 34: Article 19 (Freedoms of Opinion and Expression), Human Rights Committee 102nd session, UN Doc CCPR/C/GC/34, September 12, 2011.

opinion, so-called forum internum, is absolute and non-derogable.[18] Article 19 of the Universal Declaration of Human Rights as well as the International Covenant on Civil and Political Rights protect these absolute rights from any unjustified restrictions and interferences. The right to form one's opinion is an essential part of the freedom of opinion. In the words of the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, "any involuntary disclosure of opinions is prohibited and mental autonomy is affirmed".[19]

Therefore, the distinction between the internal aspect of the right to freedom of thought and freedom of opinion, i.e. the right to think or believe, and the manifestation of the right, is essential. Freedom of thought as well as freedom of opinion are absolute freedoms, enshrined in international human rights treaties, including the European Convention on Human Rights and the International Covenant on Civil and Political Rights. Due to its absolute nature, no interference with these freedoms can be justifiable. The international human rights framework includes in the scope of the freedom of thought and freedom of opinion three main elements:

1) The right to keep one's opinion private;

2) The right not to have one's opinion manipulated; and finally

3) The right not to be penalised for one's thoughts.

3.3 Right to privacy

> ***Charter of Fundamental Rights of the European Union***
>
> ***Article 7***
> *Respect for private and family life*
>
> 1. Everyone has the right to respect for his or her private and family life, home, and communications

Intertwined with the right to freedom of thought, conscience, and religion is the right to privacy. This fundamental right is also protected under the European Convention on Human Rights and in the International Covenant on Civil and Political Rights.

Internet users might not want to share their opinions and beliefs; however, the current practice of surveillance-based advertising enabled by massive data collection incentivises the disclosure of such information and allows internet intermediaries to infer these protected attributes. Ethnicity, sexual orientation,

18 Office and the High Commissioner for Human Rights, *CCPR General Comment No. 22: Article 188 (Freedom of Thought, Conscience and Religion)*, 1993.

19 Irene Kahn, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, *Disinformation and freedom of opinion and expression*, 2021.

political affiliation, or religious belief can be determined by a proxy, e.g., through geolocation services, app use, or internet browsing habits. This profiling can happen even without the advertiser's knowledge, as it is fueled by the optimisation of advertising algorithms. This can in turn lead to automated discrimination as users are targeted based on these sensitive categories or, on the other hand, face categorical exclusion.[20]

In its opinion on the Digital Services Act, the European Data Protection Supervisor (EDPS) highlighted the importance of transparency measures for users and accountability provisions for advertisers.[21] To better protect people's fundamental rights, both the EDPS and the European Data Protection Board (EDPB) urge the EU legislators to consider a phase-out leading to a prohibition of targeted advertising on the basis of pervasive tracking.[22]

3.4 Data protection

> ***Charter of Fundamental Rights of the European Union***
>
> ***Article 8***
> ***Protection of personal data***
>
> 1. Everyone has the right to the protection of personal data concerning him or her.
>
> 2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
>
> 3. Compliance with these rules shall be subject to control by an independent authority.

At its core, data protection is about preserving a fundamental right that is reflected in the Charter of Fundamental Rights of the European Union, Council of Europe Convention 108, and other international agreements and national constitutions. The adequate enforcement of the General Data Protection Regulation in the European Union could drastically reduce some of the worst practices of the surveillance-based industry.[23]

20 Forbrukerrådet, *Time to ban surveillance-based advertising*, June 2021.
21 European Data Protection Supervisor, *Opinion 1/2021 on the Proposal for a Digital Services Act*, February 10, 2021.
22 EDPB, *Statement on the Digital Services Package and Data Strategy*, November 18, 2021.
23 Access Now, *Three Years Under GDPR*, 2021.

IV. How Human Rights Abuse Happens In Practice

In our 2018 joint report, *Informing the "Disinformation" Debate*, we underlined the importance of distinguishing between the manipulative business models of large online platforms, and their role and economic interest in the spreading of dis/misinformation, and state-led "hybrid threats" such as cyber attacks and disinformation campaigns.[24] Because large online platforms with economic dominance harvest an unprecedented amount of personal data, they are able to boost the engagement of their users and derive profit by prioritising or quantifying the popularity of certain types of sensational content, including disinformation. Their dominant position also enables them to control the online public sphere, while deepening huge power asymmetries between them and their users. In the hands of major players, the acts of content moderation and content curation have become a commodity from which platforms generate profit.[25]

This report identifies two main intrusive techniques that are a driving fuel of large platforms' business model:

1. Surveillance-based advertisement, or in other words, digital advertising that is targeted to individuals, through tracking and profiling based on personal data. Surveillance-based advertisement has significantly contributed to the exploitation of people's particular characteristics to increase the persuasiveness of a message and therefore, negatively impacts their absolute freedom to form an opinion and their thought processes.

2. Amplification of potentially harmful but legal content, including disinformation, via content recommender systems and news recommenders, that contributes to the polarisation of opinions and attitudes online. Since controversial issues in particular generate user engagement, these issues are more likely to be highly ranked by algorithms and thereby more likely to be visible to a larger audience on social media.[26] Content recommendation is crucial for the growth and dominance of large platforms, and lies at the heart of their business models. In the words of Tarlton

24 Access Now, Civil Liberties Union for Europe, and European Digital Rights (EDRi), *Informing the "Disinformation" Debate*, October18, 2018.

25 Tarleton Gillespie. *Custodians of the internet*, 2018.

26 Ibid. 24

> Gillespie, recommendation systems are "a key logic governing the flows of information on which we depend".[27]

This report dedicates specific focus to political advertising and personalisation of politically sponsored content online in order to inform the recently launched proposed Regulation on the transparency and targeting of political advertising.[28]

4.1 Amplification of disinformation by algorithmic curation

Personalisation and content recommender systems used by large internet intermediaries increasingly raise concerns over potentially negative consequences for diversity, the quality of public discourse, and privacy. The algorithmic filtering and adaptation of online content to speculated personal preferences and interests is often associated with a decrease in the diversity of information to which users are exposed.

Algorithmically driven content curation is a powerful tool that can profoundly influence the thought process and opinions of online users. As a consequence, amplification of disinformation and other categories of potentially harmful content undermines users' ability to arrive at well-informed opinions and makes them more vulnerable to manipulative interference by external actors. The business models of very large online platforms are built upon intrusive data practices and a persuasion architecture that can be used to manipulate and persuade people at a large scale. Personalisation of content may have a significant effect on the cognitive autonomy of individuals and interfere with their right to form an opinion.

Content recommender systems are also used by media and news sites. The main purpose of news recommenders is to filter large amounts of information online. Based on the scholarly work in this area,[29] there are three basic types of recommender systems used by news media sites. The first category consists of algorithms that create personalised recommendations based on user data (for example, their reading history, personal preferences, etc.), also known as content-based recommenders. Second, so-called collaborative filtering systems calculate recommendations on the basis of what friends or similar users to a specific user in question have liked, shared, or purchased online. These types of recommenders directly rely on users' data. And finally, the third category is a hybrid model of all previous categories. However, especially in the context of news media sites, it is essential to distinguish between self-selected recommendations when

27 Ibid. 24.

28 European Commission, *Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the transparency and targeting of political advertising*, November 25, 2021.

29 Natali Helberger, *On the Democratic Role of News Recommenders*, 2019.

users determine their own selection criteria based on their preferences, and pre-selected recommendations, i.e., those a news site selects based on data voluntarily provided or inferred from their users.

Content recommender systems rank the user-generated content using aggregated patterns derived from the behavioral data of users, with the goal of predicting what content users would like to see. This data include logs of seemingly private clicking and browsing behaviour online, while users usually remain unaware of ongoing algorithmic content curation. Logged data can be used in either anonymised aggregated form, to rank recommendations for all users, or in non-anonymous form to create personalised recommendations tailored to a specific user.[30] In order to understand the potential negative implications of these systems, they need to be considered in relation to the surveillance-based advertisement model of which they are integral. In the words of the European Parliament, "a platform that optimises for ad revenue has reason to prioritise 'content based on addressing emotions, often giving rise to sensation in news feed and recommendation systems'".[31]

Even though personalisation and content recommendation systems used by large online platforms increasingly raise a number of concerns, it can also be valuable to users when it is used to refine search and speed up the retrieval of information. Personalised search can help users find their way through the digital abundance of online information and, as such, users can find it a potentially very useful tool, especially if it is user-driven, or if there is an enhanced possibility for users to drive and control such search. However, there needs to be a clear line drawn between the use of personalisation for search (active) and for targeting (passive). Research demonstrates that when data is initially input by users (active personalisation), it tends to produce a greater diversity of information, whereas personalisation that is selected by systems (passive personalisation) could tend to have "a negative effect on knowledge production" among other things, exacerbating the so-called filter bubble effect or amplification of potentially harmful content. Passive personalisation gives online platforms the power both to decide what news and information is displayed at the top of the search box and also which advertisements will come with that material when we click on it. Provocative material usually gets the most clicks and earns the advertisers the highest revenues, and therefore could be intentionally positioned in a prioritised manner.[32]

30 Daphne Keller, *Amplification and Its Discontents: Why regulating the reach of online content is hard*, 2021.

31 Committee on Legal Affairs, *Draft Report with recommendations to the Commission on a Digital Services Act: adapting commercial and civil law rules for commercial entities operating online (2020/2019(INL))*, April 22, 2020.

32 The functioning of the engagement-based business model on Facebook has been well documented in the recent revelations by Facebook's whistleblower Frances Haugen. Dan Milmo, *Facebook revelations: what is in cache of internal documents?*, October 25, 2021.

In order to move users away from a merely passive role to a more active and empowered position in the online ecosystem, policy makers should aim at creating a decentralised environment of content recommender systems.[33] Current content curation performed by very large online platforms is provided in a bundle, i.e., users have zero choice and have to accept the whole package. In order to deliver a meaningful change to the prevailing status quo of platforms' dominance, hosting of user-generated content needs to be separated from content curation, enabling creation of alternative content recommender systems by verified third parties.

For this diversified environment to flourish, third-party recommender systems must be able to operate on social media platforms, meaning they need to be interoperable with them. Such interoperability requirements will empower citizens to choose the recommender system that aligns most with their values and interests, while also spurring innovation and competition among providers, and stimulating media diversity and information pluralism. In this scenario, very large online platforms will no longer have immense power over our information diet, shifting control back to the people.[34]

4.2 Surveillance-based advertisement: Ad tech as the financial driver for amplification of potentially harmful content

Using online services and connecting with others using online platforms often means being, directly or indirectly, subject to profiling and targeting by the surveillance-based advertising industry, also known as the ad tech industry.[35] Especially since the mid-2000s, online advertising has become the most important economic foundation of dominant online services. And, even when a service is not based on targeting ads, many websites and apps use some sort of tracking and hand over personal information to third parties such as data brokers or analytics services.

Companies such as Facebook/Meta employ surveillance advertising by using user data as the basis for decisions about the advertisements that users see in their news feeds, based on what will likely appeal to them and they will subsequently engage with and click on. This type of data manipulation reinforces the need for the ePrivacy Regulation to enter into force as soon as possible as a means of changing the balance of incentives for companies away from a model that relies on sensationalism

33 Panoptykon, *Big Tech platforms are hurting us. 50 organisations urge the EU to #fixalgorithms,* September 22, 2021.
34 European Digital Rights (EDRi), *Can the EU Digital Services Act contest the power of Big Tech's algorithms?*, 2021.
35 EDRi, *Targeted Online: An industry broken by design and by default*, 2021.

and shock to artificially boost user engagement. What is more, the current surveillance advertising industry financially contributes to the creation and spread of disinformation on the internet. A *New York Times* investigation found that some of the most harmful false and misleading news stories that circulated widely during the 2016 US presidential election had been created and spread with the explicit goal of earning advertising revenue through Google's surveillance ad network.[36]

Even though not all disinformation, hate speech, and polarisation are caused or intensified by surveillance advertising,[37] the industry's role has exponentially gained in importance as it acquired more and more personal data about internet users around the world. An analysis by Ranking Digital Rights documented[38] how the so-called attention economy, built through surveillance-based advertising that is fed with personal data, intensified and exacerbated these problems worldwide.

Evidence seems abundant and unambiguous. According to the Global Disinformation Index,[39] private companies and political actors have fueled extremist and disinformation websites with at least $235 million in revenue generated annually from ads which run on those websites. *News Guard* reported that top brands are sending $2.6 billion to misinformation websites each year.[40] Surveillance-based advertising is a key funder of online hate and disinformation which can disrupt elections, incite violence, and prevent us from effectively tackling climate change.

It has been documented[41] how platforms like YouTube and Facebook/Meta, that run recommendation algorithms in their surveillance-advertising platforms, have given priority to disinformation in their platforms. Once again, the focus on "engagement" rates (time spent on their platforms) made it more profitable for those platforms, as more time spent on them meant more time being exposed to advertising.

The European Parliamentary Research Service (EPRS) has called such trends in the online advertising industry "polarisation by design".[42] In essence, the more radical the content is (that is, the more extreme, provocative, or divisive),

36 MIT Technology Review, *How Facebook and Google fund global misinformation*, 2021.

37 Glenn Greenwald, *The CIA's Murderous Practices, Disinformation Campaigns, and Interference in Other Countries Still Shape the World Order and U.S. Politics*, 2020.

38 Nathalie Maréchal and Ellery Roberts Biddle, *It's Not Just the Content, It's the Business Model: Democracy's Online Speech Challenge*, 2020.

39 Global Disinformation Index, *Cutting the Funding of Disinformation: The Ad-Tech Solution*, 2019.

40 News Guard, *Special Report: Top brands are sending $2.6 billion to misinformation websites each year*, 2021.

41 Avaaz, *YouTube and climate disinformation*, 2020. and Avaaz, *Facebook and coronavirus misinformation*, 2020.

42 European Parliamentary Research Service, P*olarisation and the use of technology in political campaigns and communication*, 2019.

the more interaction there will be between users and that given content. Whether it is because people disagree with it, or they simply find it provocative, people tend to spend more time engaging with that type of content. The content, broadly speaking, does not get more views based on its quality or the depth of the research on a given topic, but because it is successful in generating interest and exposing people to ads. Because of these business models, true investigative journalism rarely benefits, as publishers are incentivised to use snarky headlines or clickbait traps to provide more "provocative" (and profitable) content.

Furthermore, bad faith actors can use surveillance-based advertising platforms to create division and disrupt political discourse. The Facebook Cambridge Analytica scandal and the Brexit campaign have shown that surveillance-based ads, especially when combined with other types of abuses of personal data, can easily be used to micro-target people with false information in order to try to shift their opinions outside the public discourse and regardless of the public opposition to those practices.[43] Facebook in particular enables precision-targeted political and other messages, thanks to its access to behavioural data and algorithms, both treated by the platform as its "property".[44]

Surveillance-based advertising can discriminate against and exclude certain groups from accessing information, which can intensify marginalisation and social exclusion. Advertisements about employment, housing, or elections can be hidden from certain people, based on age, gender, location, or more sensitive data, like ethnicity, political and sexual orientation, or browsing behavior. This was demonstrated in a study[45] by investigative journalists who published housing advertisements and, using Facebook's targeting tools, excluded certain groups, such as Black Americans, mothers of high school kids, or people interested in wheelchair ramps.[46]

Even the advertising industry seems to be waking up to the risks this system poses. The Norwegian branch of the World Federation of Advertisers, AFNO, published an article[47] arguing that more and more businesses realise that self-regulation does not work and that new regulation is needed, and suggested a solution: the use of contextual ads or "new alternative advertising systems where challenges related to transparency, ownership, and data security

43 European Parliamentary Research Service, *Polarisation and the use of technology in political campaigns and communication*, 2019.

44 Ian Bogost and Alexis Madrigal, *How Facebook Works for Trump*, 2020.

45 Julia Angwin et al, *Facebook (Still) Letting Housing Advertisers Exclude Users by Race, ProPublica*, November 21, 2017.

46 Jascha Galaski and Eva Simon, *Solutions for Regulating Targeted Political Advertising on Online Platforms*, May 3, 2021.

47 Jan Morten Drange, *Digital målrettet reklame under angrep!*, 2021.

are solved". In fact, it is important to note that the business-model of online platforms was not always reliant on the use of targeted ads and that alternative models can exist.

Finally, recent research has shown the traumatic experiences that surveillance advertising might cause.[48] If the illegal collection and access to citizens' data were stopped, micro-targeted disinformation campaigns would lose much of their alleged effectiveness and threat potential. As is already clear, weak enforcement of data protection rules and a lack of updated privacy legislation not only impacts user privacy and choice, but also lead to the constant monitoring, profiling, and "nudging" of people in line with the interests of those actors who can afford to invest in surveillance-based advertising.

4.3 Political advertisement

We learned from Brexit and the 2016 US presidential election that targeted political advertisements online and offline have the potential to significantly reduce fairness and influence the outcome of a vote. Political advertisers can target people based on their behavioural data that is collected and made available by online platforms. Political advertisers can use the data to segment groups of people susceptible to being convinced by a given message and send those people highly personalised appeals to support a particular candidate or policy proposal. Targeting techniques are convenient for advertisers. They make it easier to find very narrow segments of the population and can be used to mislead, manipulate, or demobilise voters by delivering different messages to different groups of voters, creating ideological echo chambers for them, and limiting their right to get access to information. Moreover, it is easy to flood these echo chambers with tailored disinformation or extremist content that can polarise people.

As civil society group Panoptykon Foundation puts it in their report on political ads,[49] "surveillance-based business models rely on constant data collection and profiling, the platforms aim to maximise the time users spend on the platform by using ranking algorithms that promote content that is more engaging. It has a very negative impact on public health, especially for young people who become addicted to social media.[50] However, it is also detrimental to the quality of media and gives rise to clickbait and fake news, given that human psychology reacts more strongly to emotional or sensationalist messages". Limitations on targeting methods would force political actors to present a consistent agenda to the general public, and would support open public debates.[51]

48 Dorota Głowacka and Karolina Iwańska, Panoptykon, *Algorithms of Trauma*, 2021.

49 Panoptykon, *Who (really) targets you?*, 2021.

50 See for example: Yubo Hou et al., *Social media addiction: Its impact, mediation, and intervention*, 2019.

51 Dorota Głowacka and Karolina Iwańska, Panoptykon, *Algorithms of Trauma*, 2021.

V. Policy recommendations addressed to the European co-legislators

Phase out advertising that is based on tracking and targeting based on personal data, including inferred data. In order to address the core of the currently toxic business model, targeting in advertising should be limited to information that people provided voluntarily, specifically, and explicitly for that purpose. Companies should not be able to collect additional information for that purpose or combine data they may have on users or so-called look-alike audiences. Companies should not track users on their platforms or elsewhere for advertising purposes. People should be able to access, review, and change what the platforms know about them when they are targeted with specific content or ads. The advertising model and content curation should be based on contextual information and only personalised based on the preferences that people provide voluntarily, without being nudged via forced "consent". In order to ensure this, "dark patterns" practices[52] must be banned, and automated signals (like "Do Not Track") and other privacy-by-design and by-default features must be imposed on browsers, websites, operating systems, hardware, and apps to guarantee people's security and privacy.

In the transition to phasing out surveillance-based advertising, limit targeting methods to the minimum and provide transparency on the current targeting methods. Regulators should limit the targeting methods that online platforms make available. Targeting methods based on behavioural data, both observed (e.g. what sort of content users like and share) or inferred (assumptions that algorithms make about users' preferences based on surveillance of people's online activity and the building of profiles, which can be discriminatory) should be fully prohibited. This limitation of targeting criteria would reduce the possibility that political actors tailor different messages to different groups of people and manipulate or even mislead the electorate. Instead, we believe that methods not based on surveillance, such as contextual advertising,[53] offer the best way forward. Furthermore, online platforms should be subject to *meaningful and qualitative transparency obligations supported by proper independent oversight.* This includes mandatory disclaimers on all political and issue-based advertisements, including detailed information on why, how, and by whom advertisement recipients are targeted, as well as mandatory archives with detailed information on paid content. The archive should contain, among other things,

52 See more on dark patterns in Norwegian Consumer Council, *Out of control*, January 14, 2020.

53 Johnny Ryan, (*Six Months of Data): lessons for growing publisher revenue by removing 3rd party tracking*, July 24, 2020.

the advertisement's content, the targeting criteria used to reach out to online platform users, the amount spent, the time it started and the time it stopped, and the performance of the advertisement. The archive must be publicly available, easy to navigate, and designed to facilitate research and analysis. Public access to information related to direct and indirect payments or any other remuneration received to display advertisement must be ensured. Civil society, independent researchers, relevant authorities, national electoral commissions, other public authorities, and regulatory bodies should be able to monitor and evaluate political advertising and better understand its impact on democracy and fundamental rights.

Mandate accountability for platforms' delivery algorithms to help ensure proper oversight. As a part of meaningful transparency for people, it is necessary to ensure that content-recommender models are being adequately explained to users. Explanation of the family of models, input data, performance metrics, and how the model was tested should be communicated to users in tangible and comprehensible language. Such an explanation, with sufficient technical details, will allow users, scientists, regulators, and NGOs to contest the algorithmic decision-making and/or to opt-out. The right to object to the use of automated decision-making systems should apply even if a human is involved in the process. Trade-secrets should not be invoked as an excuse to not disclose how algorithms and advertising works.

Ensure a strong enforcement of the General Data Protection Regulation and the adoption of a strengthened ePrivacy Regulation to eliminate intrusive targeting techniques and limit the spread of disinformation. The national Data Protection Authorities (DPAs) must properly apply and enforce the GDPR. The GDPR safeguards EU residents' data protection rights and prevents the misuse of their personal data, including for targeting purposes. Targeting on the basis of sensitive and protected characteristics is prohibited and must be enforced. This includes grounds of discrimination under the EU Charter: race, disability, social origin, and others. These practices have been shown to be highly problematic and lead to discrimination of marginalised groups.[54] EU regulators and co-legislators should work to eliminate dark patterns that online platforms use to trick users into disclosing their data, such as "I agree" buttons that users click to get rid of annoying pop-ups or banners. Well-informed, specific, and explicit consent on behalf of the user is needed prior to processing personal data for targeted advertising. Even though the GDPR provides solid ground for valid consent requirements, the Digital Services Act, the Digital Markets Act and the relevant upcoming proposal for targeted political advertising are also addressing these questions. The legislators should ensure that there are no contradictions and redundancies

54 European Digital Rights (EDRi), *How online ads discriminate: Unequal harms of online advertising in Europe,* June 2021

between co-existing pieces of legislation. Specifically, the GDPR and the ePrivacy Directive (and the future ePrivacy Regulation) should remain the baseline to build on. The role of the DSA and DMA could complement or clarify the aspects that neither of these pieces of legislation have tackled yet, including clarifying specific bans of highly intrusive practices such as surveillance-based advertising and dark patterns.

Establish minimum safeguards for users' default settings to require an "opt-in" to personalised content recommendations systems rather than the current default "opt-out" in the ongoing discussion on key digital policies (DSA, DMA, ePrivacy Regulation). Platforms should design "consent" and privacy policies in a way that facilitates informed choice for users and is compliant with data protection laws. Users have to be able to exercise control over recommendation systems that can be secured by an "opt-in" mechanism. Making content recommender systems available via "opt-in" would be a desirable mechanism because even those users who are less aware of how these systems operate will not be treated less favourably. This means that such content recommender systems should be off by default and only activated by the users who should not be forced to do so by prompt. Those users who decide to receive content recommendations should be able to:

- Exclude certain content from their recommendations;
- Exclude certain sources of content from their recommendations; and
- Ask for profiles to be deleted and access the service even when refusing to use content recommendations, to ensure the opt-in is meaningful. Users should be able to do so in an easy and free manner, and at any time they wish.

Complement the protection for human rights online afforded under the GDPR through the DSA and ePrivacy Regulation. The draft ePrivacy Regulation and the draft Digital Services Act offer the possibility to complement general rules provided by the GDPR that apply to the context of targeted ads. In addition, the Commission should urge the Member States to provide DPAs with the funds necessary for the tasks they are expected to undertake in the protection of the rights to privacy and data protection. We encourage the quick adoption of a strengthened ePrivacy Regulation with strong privacy by design and by default protections to help the development of better control mechanisms for users' rights.

Require mandatory Data Protection Impact Assessments and ex ante mandatory Human Rights Impact Assessments (HRIA). In fulfilling their transparency obligations, political parties, interest groups, and platforms should be required to conduct and publish Data Protection Impact Assessments and a Human Rights Impact Assessment relating to online political campaigns hosted on relevant platforms. We advocate introducing a HRIA that analyses the effects that business activities have

on users. An HRIA[55] follows a human rights-based approach, which integrates human rights principles such as personal data protection, non-discrimination, and freedom to access information into the assessment process.

Empower people. There is a severe power imbalance between online platforms and users. Users should have more control over their news feed and their personal data online. First, people should be allowed to decide whether they want to receive targeted political advertisements or not. For this to happen, and in accordance with EU data protection rules, online platforms should receive users' explicit and specific consent to use data for this purpose. Sensitive data however shall never be used for this purpose. Second, and to limit pop-up fatigue, automated signals (as mentioned above) must be used by default and be binding; furthermore, there should be rules that limit how often online platforms can ask users to opt-in and that ban dark patterns. Thirdly, companies must provide a mechanism where people can learn about companies' targeting methods, the data processed, and the rights set out in Article 15 of the GDPR. Online platforms should have 15 days to answer such requests. Finally, people should be empowered to choose their own content recommendation systems which may be based on a different logic than engagement maximisation. These steps should help with the phasing out of the use of targeted ads online towards a more sustainable business model for all online platforms, media, and for the protection of users' rights.

Support digital and media literacy. To minimise the impact of disinformation, it is important to continue educating about critical thinking, and to focus on life-long learning about the use of digital technology and assessment of reliable sources of information. Such a project, conducted at the national level, and potentially financed by the Commission, should not only focus on the younger generation but also the elderly .

Invest in and support media pluralism and freedom. Encourage quality journalism, media collaborations, and independent media outlets, and support policy for breaking up the media conglomerates that dominate public discourse.

Create and foster a sense of belonging and community online for people leaving hate groups or abandoning conspiracy theories. Encourage and explore models to support independent fact checking or community-based fact checking, such as those used by the Wikimedia Foundation in its projects.

Strengthen the control of public funds that fuel state propaganda. Research in Slovenia[56] shows that European Union and other public funds are being used to spread state or party propaganda. Therefore, we suggest considering

55 The Danish Institute for Human Rights, *Human rights impact assessment guidance and toolbox*, August 25, 2020.

56 Domen Savič, *Spreading propaganda and disinformation using public funds The case of Slovenia as a challenge for EU democracy*, July, 2021.

recommendations such as: ensuring transparent reporting on advertising campaigns; ensuring the independence of the advertising industry from state intervention; creating an independent advertising regulatory agency; establishing a special framework for ad fraud investigation; allocating special EU funding for ad fraud investigative projects; and exploring ways to reduce the power of Big Media and its influence on public authorities.

VI. Conclusion

The European Union is standing at a crossroads. With upcoming regulatory efforts to correct the huge power asymmetry between very large online platforms and the people who use them, it has a chance to establish clear rules that will prevent manipulation of users and ongoing human rights abuse. While recently proposed sectoral solutions, such as the Regulation on political advertising, finally recognise the amplification of polarising content and pervasive targeting techniques as a common denominator in fundamental rights violations, singling out a tiny piece of the disinformation puzzle will not deliver the sustainable solutions that can revolutionise the online ecosystem.

Therefore, in our view, only a horizontal regulatory framework can eliminate the negative impact of systems and strategies deployed by very large online platforms that enable the spread of disinformation and other forms of potentially harmful content, in a compliance with fundamental rights standards. Currently negotiated horizontal frameworks within the Digital Services Act and Digital Markets Act proposals offer an extraordinary opportunity to effectively minimise the negative impact of amplification, personalisation, and pervasive targeting techniques through proper enforcement of existing rules. Targeted advertising is at the core of disseminating disinformation. In order to prevent the spread of disinformation elaborated in "fake news" factories, we need a new model of human rights-centric platform governance that consists of effective enforcement of the General Data Protection Regulation; introducing mandatory risk assessments such as a Human Rights Impact Assessment; swift adoption of the proposed e-Privacy Regulation; and a fundamental rights-oriented DSA and DMA.

VII. Glossary of terms

Ad tech / Online tracking industry / Behavioural advertising / Micro-targeted advertising: This is the industry that collects and/or processes personal data for the purpose of customising ads or content. It does this by profiling people, collecting their data when they browse the internet or use apps and online services/platforms. We will use the terms interchangeably, even though we may be referring to different services, companies, or business models.

Automated decision-making: This refers to algorithms used across a variety of domains, from simplistic models that help online service providers to carry out operations on behalf of their users to more complex profiling algorithms that filter systems for personalised content. Automated, algorithmic decision-making is usually difficult to predict for a human being and its logic will be difficult to explain after the fact.

Content recommender systems: "Recommender system" is a term that describes various technologies that help users filter and retrieve information. The information processed by a system covers a wide range of items, ranging from songs to books, movies, news articles, and more. There are two filter strategies available to provide users with item recommendations:

1. Content-based filtering: Users get item recommendations based on their preferences. E.g., if someone likes classical music or news about their favorite sports team, then the recommender system will look at the content of the item and only provide the user with the items that align with their interests.

2. Collaborative filtering: Users get item recommendations based on people they are closely associated with. E.g., when a user is reading news, a system recommends articles a friend has shared/read, or when a user does online shopping, the system recommends articles that people with a similar shopping history have purchased.

The filtering method describes what type of data is used by the recommender system. It is important to note, however, that the two filtering techniques are not mutually exclusive. More and more so-called hybrid recommender systems are used that combine the two approaches (e.g., recommending a news article a friend has liked, but only if it covers a certain topic). The goal of deploying a recommender system is content personalisation; i.e., a user of a platform or service gets recommended content that is custom-tailored to their profile (personal interests and relation to other users). Accordingly, the videos, search results, news articles, or any other type of content that is displayed to the user can be unique to their experience and differs from what other users see. It is possible for platform or service owners to enrich their user data by purchasing

additional information from third parties (i.e., other platforms or services).

Disinformation: There is no universally agreed definition of disinformation but we build on the agreed definitions by the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression[57] to define it as statements which are known or reasonably should be known to be false that are disseminated intentionally to cause serious social harm. Disinformation misleads the population, and as a side effect, interferes with the public's right to know and the right of individuals to seek, receive, and impart information.

Misinformation: This refers to false information disseminated unknowingly.

Online platform: We use the term in this paper to cover a number of very different services that have in common the provision of goods or services to the public online, such as social media platforms, e-commerce businesses, search engines, and apps. In our discussion, we refer specifically to those platforms using targeting techniques to personalise ads or display content.

State-sponsored propaganda: Statements which demonstrate a reckless disregard for verifiable information and that are sponsored and/or funded, directly or indirectly, by State authorities.

Surveillance-based advertising: A blanket term for digital advertising that is targeted to individuals, usually through tracking and profiling based on personal data.

57 Irene Khan, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, *Disinformation and freedom of opinion and expression*, point 2, April 13, 2021.

Access Now defends and extends the digital rights of users at risk around the world. By combining direct technical support, comprehensive policy engagement, global advocacy, grassroots grantmaking, legal interventions, and convenings such as RightsCon, we fight for human rights in the digital age.
accessnow.org

Eliska Pirkova, Access Now, eliska@accessnow.org

The Civil Liberties Union for Europe (Liberties) is a non-governmental organisation promoting and protecting the civil liberties of everyone in the European Union. We are headquartered in Berlin and have a presence in Brussels. Liberties is built on a network of national civil liberties NGOs from across the EU. Unless otherwise indicated, the opinions expressed by Liberties do not necessarily constitute the views of our member organisations.
liberties.eu

Eva Simon, Civil Liberties Union for Europe, eva.simon@liberties.eu

European Digital Rights (EDRi) is an association of civil and human rights organisations from across Europe. We defend your rights and freedoms in the digital environment.
edri.org

Diego Naranjo, European Digital Rights, diego.naranjo@edri.org