

This publication is part of the project “Spotlight on Artificial Intelligence and Freedom of Expression” (#SAIFE).

The views, findings, interpretations, recommendations and conclusions expressed herein are those of the authors and do not necessarily represent the official position of the OSCE and/or its participating States.

© 2021, Office of the Representative on Freedom of the Media
Organization for Security and Co-operation in Europe (OSCE)

6a Wallnerstrasse
1010 Vienna, Austria
Phone +43-1-514-36-68-00
e-mail: pm-fom@osce.org
<https://www.osce.org/fom/ai-free-speech>

ISBN: 978-92-9234-749-9

Electronic copy available at: <https://ssrn.com/abstract=4060166>

AI in Content Curation



AI in Content Curation and Media Pluralism

This part focuses on the use of AI in content curation, addressing the impact of data-driven content recommender systems on diversity and media pluralism. This part and the next one highlighting shortcomings of AI-based content curation and targeted advertising provide human rights-centred recommendations to prevent the negative impact of AI tools in content curation on the right to freedom of opinion and expression.

1. Defining the scope of content curation's impact on media pluralism

1.1 The relevance of algorithmic content curation and data-driven recommendation systems to media pluralism and diversity

Diversity and media pluralism are core democratic principles whose quality is impacted by the rise of dominant internet intermediaries and their influence over public discourse. Internet intermediaries, in particular social media platforms, have become an important source of, access point for and key distributor of information, including news content. Information dissemination and, increasingly, aggregation occurs primarily through algorithmic content curation²⁶ and recommender systems. Using optimisation and analysis of human and non-human agents, these systems “deliver” personalised content customised to individual profiles, resulting in the type and amount of content to which each individual is exposed. Content recommender systems, which rank content to determine what is presented to individual users, impact individuals’ freedom to seek and impart information, as well as the overall information landscape and media freedom. The design of recommender systems significantly affects what is seen online, and what remains hidden—and for whom. The

²⁶ Content curation can be understood as a set of algorithmic and human-driven processes that support the distribution of content to audiences, such as content ranking or editorial data analysis. See at: B. Bukovska et al, Spotlight on Artificial Intelligence and Freedom of Expression #SAIFE (2020), p.19.

process of algorithmic curation is underpinned²⁷ by the values and goals of the algorithm's creator,²⁸ socio-technical factors, self-regulation (Terms of Service, for example) and state regulation. Given how ubiquitous online content has become, and the significance it has in shaping opinion and decision-making, the pivotal question arises: Where does responsibility lie in defining and implementing policy to prioritise and codify media pluralism and diversity²⁹ in the era of digital information?

The part of the report provides a conceptual summary of key algorithmic-curation processes and their transformative impact on media pluralism. It further provides a set of recommendations for OSCE participating States on a human-rights-centred approach to algorithmic content curation. As such, this part focuses on the impact that algorithmic content curation and data-driven recommendation systems have on media pluralism and diversity in democratic societies—and the state's role to act as the ultimate guarantor of the human right to freedom of expression and to ensure an enabling environment for this expression.

1.2 Incongruence of algorithmic content curation and freedom of expression

The ability to filter, prioritise and engage with online content based on personal preferences and interests is often at odds with the individual agency to seek, receive and impart diverse information.³⁰ As a basic principle, internet intermediaries typically prioritise and display content to an individual based on the system's prediction that the individual is

²⁷ K. Klonick, The New Governors: The People, Rules, and Processes Governing Online Speech, *The Harvard Law Review*, p.1664.

²⁸ Radsch, Courtney. "Digital Information Access." In *A New Global Agenda: Priorities, Practices, and Pathways of the International Community*, edited by D. Ayton-Shenker, 72–83. Rowman & Littlefield Publishers, 2018. <https://books.google.com/books?id=tyJL-DwAAQBAJ>.

²⁹ On exposure see: Philip M Napoli, "Rethinking Program Diversity Assessment: An Audience-Centered Approach" (1997) 10 *Journal of Media Economics* 59–74.; N. Helberger & M. Wojcieszak (2018). Exposure Diversity. In P. M. Napoli (Ed.), *Mediated Communication* (pp. 535–560). (*Handbooks of Communication Science*; Vol. 7). De Gruyter Mouton. <https://doi.org/10.1515/9783110481129-029>.

³⁰ P. Leersen, The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems, *European Journal of Law and Technology* (2020), p.12.

likely to engage with that content. Similar to systems for personalised and behaviour-based advertisements, content recommender systems thus extensively collect data of users (and non-users) to create digital profiles, assess similarities and make inferences based on this data.

Many online platforms' business model,³¹ which prioritises engagement and profit over a human rights-centred approach, can and does result in exploitative and intrusive data practices, the spread of mis/disinformation and algorithmic feedback loops.³² It has been proven to have a negative influence on content plurality, especially regarding content created by or for marginalised communities. The model perpetuates information gaps³³ and constitutes obstacles to advocacy, thereby recreating and bolstering structural societal inequality. There is also evidence that suggests that the process of content moderation benefits those groups already dominating online spaces and narratives over marginalised groups, information and narratives.³⁴ Moreover, algorithm-driven content discovery (e.g. search engines) have been found to reinforce racism by suggesting discriminatory search phrases and discrepancies, particularly along racial, language and gendered lines, in depictions of members of marginalised communities.³⁵

For the most part, algorithmic content curation and recommender systems are based on intermediaries' own (internal) rules, interests and assumptions, rather than democratic or public interest values.³⁶

³¹ See: Ranking Digital Rights, It's the Business Model: How Big Tech's Profit Machine is Distorting the Public Sphere and Threatening Democracy (2020).

³² Bodó, B., Helberger, N., Eskens, S., & Möller, J, Interested in diversity: The role of user attitudes, algorithmic feedback loops, and policy in news personalization. Digital Journalism (2019), p.219.

³³ A. Causevic and A. Sengupta, Whose Knowledge Is Online? Practices of Epistemic Justice for a Digital New Deal, IT for Change (2020), Retrieved from: <https://itforchange.net/digital-new-deal/2020/10/30/whose-knowledge-is-online-practices-of-epistemic-justice-for-a-digital-new-deal/>.

³⁴ B. Marshall, Algorithmic misogynoir in content moderation practice, Heinrich-Böll-Stiftung (2021), p.7,11. See also: M. E. Mazzoli and D. Tambini, Prioritisation uncovered: The Discoverability of Public Interest Content Online. Council of Europe (2020), p. 44.

³⁵ Safiya Umoja Noble, Algorithms of Oppression How Search Engines Reinforce Racism, NYU Press (2018).

³⁶ C. Radsch. "Digital Information Access." In A New Global Agenda: Priorities, Practices, and Pathways of the International Community, edited by D. Ayton-Shenker, 72–83. Rowman & Littlefield Publishers, 2018. <https://books.google.com/books?id=tyJJDwAAQcBAJ>.

Content recommendation is crucial for the growth and dominance of large internet intermediaries, and it lies at the heart of their business models. As recommender systems are "a key logic governing the flows of information on which we depend",³⁷ internet intermediaries are enabled to act as gatekeepers of information and knowledge. This has broader implications for public interest, representation and power (in)equality, on- and offline.³⁸ Intermediaries' recommender systems have significantly reconfigured the logic of public communication, including access to news, critical information and overall content in the public interest. Thus, their recommender systems significantly restrict equal access to, and of, journalists and media outlets, while pressuring professional journalism due to the outflow of advertising money to intermediaries. Recent research findings on algorithmic prioritisation, defined as "the range of design and algorithmic decisions that result in prominence and discoverability of content"³⁹ reveal the potential for the polarisation of opinions and attitudes online. For instance, an important factor in content prioritisation processes is individual political predisposition and/or affiliation. Prioritisation can, therefore, reinforce and perpetuate polarisation of opinions and attitudes online, especially among those users at the edges of the political spectrum who likely already consume a predominance of affiliated content.⁴⁰ It has also been shown that "some groups in society are more prone to selective exposure than others".⁴¹

³⁷ T. Gillespie (2018). Custodians of the internet. Retrieved from https://www.researchgate.net/publication/327186182_Custodians_of_the_internet_Platforms_content_moderation_and_the_hidden_decisions_that_shape_social_media

³⁸ P. Leerssen, The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems. Retrieved from file:///Users/eliskapirkova/Downloads/Leerssen%20EJLT_corr.pdf.

³⁹ M.E. Mazzoli and D. Tambini. Prioritisation uncovered: The Discoverability of Public Interest Content Online. Council of Europe (2020), p.12.

⁴⁰ B. Stark, D. Stegmann, Are Algorithms a Threat to Democracy? The Rise of Intermediaries: A Challenge for Public Discourse. Retrieved from <https://algorithmwatch.org/wp-content/uploads/2020/05/Governing-Platforms-communications-study-Stark-May-2020 -AlgorithmWatch.pdf>

⁴¹ B. Bodó, N. Helberger, S. Eskens & J. Möller, Interested in diversity: The role of user attitudes, algorithmic feedback loops, and policy in news personalization. Digital Journalism (2019), p.15.

While active personalisation based on user input tends to produce a greater diversity of information, passive personalisation based on algorithmic content selection tends to exacerbate the so-called filter bubble effect.⁴²

Bias and discrimination, including gender-based discrimination, in data-supported algorithmic decision-making can occur for several reasons and at many levels in content curation systems, and they can be difficult to detect and mitigate. It has been suggested that the exclusion of sensitive/identity-based information sufficiently protects against discrimination. Yet discrimination can and does occur, despite these “protections”, given the expansive and diverse information contained in algorithm-informing datasets. Bias in algorithms can stem from design and implementation, including unrepresentative or incomplete training data, or reliance on individual, experiential, or values-informed data that reflects historical/structural inequalities. Algorithmic bias can have a collective, disparate impact on communities, especially marginalised groups, even when there is no intention to discriminate. An exploration of both intended and unintended consequences of algorithms is thus necessary. Current public policies may not be sufficient to identify, mitigate, and remedy the impact on individuals or society at large. In addition to deliberate efforts to shape individual attention (direct manipulation), there is also a danger of unwanted and indirect biases being introduced into the algorithm through incorporation of big data at various levels of the content curation. Both direct and indirect discrimination caused by algorithms using big data are among the most pressing dangers of algorithm-driven curation processes.

The combined effect of content filtering and personalisation practically creates layers of restriction in terms of discoverability, and thus accessibility, of diverse media content. The aforementioned issues have serious implications for media pluralism, understood as a plurality of information sources (external pluralism), and of content (internal pluralism).⁴³ More specifically, and in the context of this report, media

⁴² D. Wagner, Artificial Intelligence and Disinformation as a Multilateral Policy Challenge <https://www.osce.org/files/f/documents/d/o/506702.pdf>.

⁴³ On exposure see: P. M. Napoli, “Rethinking Program Diversity Assessment: An Audience-Centred Approach” (1997) 10 Journal of Media Economics 59-74; Helberger, N., & Wojcieszak, M. (2018). Exposure Diversity. In P. M. Napoli (Ed.), Mediated Communication (pp. 535-560). (Handbooks of Communication Science; Vol. 7). De Gruyter Mouton. <https://doi.org/10.1515/978110481129-029>.

pluralism also refers to the distribution of communicative power (or “voice”) in society. A fair distribution of “voice”, as a precondition, requires the deconcentration of power and decentralisation of resources within the information ecosystem,⁴⁴ as well as support for alternative models that offer a diversity of narratives and content. It is clear that algorithmically driven content curation processes are transforming the notions of media pluralism and diversity, which are necessary for democratic, public debate and inclusive societies.

It is against this background that state and non-state actors, primarily internet intermediaries and media organisations, but also international and regional organisations, civil society representatives and academia, are called upon to adopt policies that contribute to an enabling environment for media plurality. This means enabling access, availability, discoverability and consumption of different kinds of (media) content through different mediums and via multiple channels.

2. Algorithmic content curation and data-driven recommendation systems: impact on media pluralism

2.1 Typology

A distinct source of influence, and thus communicative power, of internet intermediaries and social media companies lies in their content recommender systems, which also “lend gravitas to their role in democratic culture”.⁴⁵ In essence, a “recommender system” includes various technologies that filter, retrieve and organise information for individuals. The factors for ranking can include the level of engagement with the specific content, the type of content, when it was first shared, or how users have interacted previously with similar content. By ranking content, these systems have the potential to shape and impact individuals’ ability to form opinions.

⁴⁴ M. Moore and D. Tambini (eds) (2018) *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*. New York: Oxford University Press.

⁴⁵ K. Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, The Harvard Law Review, p.1663.

The main purpose of content recommenders is to filter large amounts of information online. This algorithmically driven process works in different ways:

- **Content-based filtering:** individuals get content recommendations based on their stated or implied preferences. For example, if someone likes classical music or news about a favourite sports team, then the recommender system will prioritise these items that align with their interests, and will, likely, encourage engagement.
- **Collaborative filtering:** individuals get content recommendations based on people with whom they are closely associated or with whom they share similarities (in demographic category, content preferences, etc.). For example, when reading news, a system recommends articles a friend has shared/read, or, when doing online shopping, the system recommends items that people with a similar shopping history have purchased.
- **Hybrid filtering:** a combination of the above-mentioned filtering and curation methods. For example, recommending a news article a friend has liked, but only if it covers a certain topic of perceived interest to the user, and combining it with a wide range of different metadata, such as an individual's location, usage history, etc.

All of these processes are based on users' data, profiles and interactions with a given platform, as well as information gleaned from the underlying ad tech architecture. The algorithm specifies the precise way in which content recommendations are generated, using content-based and collaborative filtering. The system creates a recommender strategy for how data is combined to calculate potential engagement, based on user recommendations, to satisfy optimisation criteria. Put simply, algorithmic content curation is the strategy used by a recommender system to determine how collected data can best be utilised to reach pre-defined optimisation goals.⁴⁶

46 To reach the optimisation goal, the algorithm can emphasise different ways of how to prioritise the collected data. For example, an algorithm could favour recency of a news article. Another strategy would be to look at popularity of articles as a ranking criterion of how to sort the final recommendations calculated for each user. Accuracy is another frequently used way of curating content. An accuracy-optimised approach tries to model user preference as closely as possible. They calculate recommendations that fall in line with existing user preferences. Depending on the data that was collected and the nature of the item in question, there are multiple ways of how to refine each strategy.

All large internet intermediaries, and social media platforms in particular, use so-called “open content recommender systems”.⁴⁷ These systems utilise user-generated content in the recommender’s source pool by default, but certain content items can be excluded based on, for instance, a violation of Terms of Service. To optimise engagement, these systems “personalise” the online experience by prioritising content that is assumed to be appealing and is generated through each individual’s prior engagement and behaviour. Accordingly, the videos, search results, news articles or any other type of content that is displayed to the user is unique to their experience and differs from what other users see. For this reason, among others, algorithmic content recommender systems have the potential to undermine and disrupt democratic processes.⁴⁸ They risk narrowing individuals’ exposure and access to different points of view, values and narratives, thereby threatening pluralism and diversity. This may necessitate state intervention and attention.

The algorithmic filtering and adaption of online content based on speculated personal preferences and interests decreases the exposure to a diversity of information, with potential negative effects on diversity and public discourse, as well as privacy. Content curation systems can therefore profoundly influence the information sources that form the basis for arriving at well-informed opinions, and thus the thought process of individuals. While the research is not yet conclusive, this could undermine individuals’ ability to form their opinions and make them vulnerable to manipulative interference. As the systems are built on intrusive data practices and persuasion architectures (at scale), the inevitable personalisation of content might have a significant effect on the cognitive autonomy of individuals and interfere with their right to form an opinion.

⁴⁷ As opposed to a closed recommender system that provides items to users from a limited list of options. These lists are curated by the platform owner.

⁴⁸ N. Helberger (2019) On the Democratic Role of News Recommenders. *Digital Journalism* 7(8). Routledge: 993–1012. DOI: 10.1080/21670811.2019.1623700.

2.2 Curation and prioritisation of public interest content

The methods by which online platforms curate content through recommender systems are not transparent, and they are very rarely subject to public and/or state scrutiny. When internet intermediaries incorporate diversity into recommender systems, it is typically a design choice to engage users and increase profits. Platform curated diversity has primarily been utilised to optimise financial gain, rather than to promote democratic debate, through a practice of prolonged engagement to achieve what is referred to as an optimisation goal—increased ad revenue or a higher platform/service valuation through increased traffic.⁴⁹ Put simply, business-driven content curation benchmarks have largely been established to optimise economic gain and leverage user engagement for corporate interest, rather than seeking to reflect and ensure genuinely diverse content.⁵⁰

Content recommender systems may also have unintended consequences from the perspective of broader societal objectives and can negatively shape and interfere with the absolute right to freedom of thought and opinion.

In addition, the processes of internet intermediaries' recommender systems typically exclude individual users' choice, control and agency—prerequisites to ensuring individual autonomy in seeking and imparting a variety of information and ideas. Following the public disclosure of a number of vulnerabilities of recommender systems,⁵¹ there has been increasing public and state pressure to ensure that their processes better and more meaningfully prioritise "diversified" media exposure. In particular, concerns have been raised in the context of the legality and reach of political speech, and the spread and normalisation of specific value systems as protected speech, even when content is in violation of Terms of Services or international human rights standards. Given the complete lack of information on the way in which platforms govern and prioritise

⁴⁹ In the words of a Facebook official: "Facebook is profitable only because when you add up a lot of tiny interactions worth nothing, it is suddenly worth billions of dollars.", K. Klonick, *New Governors*, p.1627.

⁵⁰ K. Klonick, *The New Governors*, p.1664.

⁵¹ A well-known and most cited case was that of the "Napalm Girl", a journalistic photo which was taken down by Facebook based on its nudity policies.

speech, it is clear that recommender systems and correlated logics of optimisation could undermine the “fair opportunity to participate”⁵² for all. At the same time, it should be recognised that human rights-based recommender systems can positively affect pluralism, for example in the contexts of authoritarianism and media capture.

It is important to define what constitutes public interest with regards to algorithmic content curation, and how it affects prioritisation of different types of content. The concept and definition of public interest content is as highly contested as the definition of diversified exposure. In principle, public interest content constitutes that information that “the public would have an interest in being informed about”.⁵³ Another way to think about public interest content is as content that is relevant to the well-being of citizens, the life of the community or the local population. Obvious examples include COVID-19 pandemic information or information related to democratic voting processes. The volume of related, and not always reliable, content has prompted intermediaries to (publicly) prioritise accuracy. Several platforms have, in a comparatively short amount of time, demonstrated their capacity to re-configure algorithm recommender systems in an effort to filter out, or label false information, and prioritise content from trusted public health authorities. The success of these efforts, or logic behind the motivation for these changes, however, remains hotly debated,⁵⁴ while a lack of transparency into the underlying data and content moderation choices made by the platforms remain a mystery. Debates aside, internet intermediaries and social media platforms—facing increasing demands from the public and states that they be held responsible for public health awareness—have shown their capacity for reflection, and for restructuring how they prioritise and rank content.⁵⁵

⁵² K. Klonick, *The New Governors*, p.1664.

⁵³ M.E. Mazzoli and D. Tambini, *Prioritisation uncovered: The Discoverability of Public Interest Content Online*. Council of Europe (2020), p. 13.

⁵⁴ M. Cinelli, *The COVID-19 social media infodemic*, *The Nature* (2020), p.10; See also: Global Disinformation Index, *Why is tech not defunding COVID-19 disinfo sites?* (2020), Retrieved from: <https://disinformationindex.org/2020/05/why-is-tech-not-defunding-covid-19-disinfo-sites>.

⁵⁵ European Commission, Joint communication to the European Parliament, the European Council, The Council, The European Economic and Social Committee and the Committee of the Regions, *Tackling COVID-19 disinformation - Getting the facts right*, JOIN(2020) 8 final, 10 June 2020., section 5.

This points to a need for greater public attention and political pressure on platforms to make transparent their recommender processes and restructure recommender systems and their optimisation goals, in order to address structural problems of our contemporary media environment. The issue goes beyond questions of content governance, and additionally concerns competition law, media ownership and concentration rules.⁵⁶ It also highlights the urgent need to prioritise media pluralism and diversity policy objectives and interventions for a more enabling digital space.

2.3 News aggregation and media plurality

News aggregators function as a central hub of online news distribution, directing readers to news items, and other content deemed (by the news aggregator) to be news. This process is predominantly carried out by algorithms, which is why news aggregators are sometimes referred to as “algorithmic gatekeepers”.⁵⁷

News aggregators often involve a tension between “algorithmic logic” and “editorial logic”.⁵⁸ “Algorithmic logic” significantly impacts diversity as well as political discourse by prioritising novelty, for example, over other criteria of newsworthiness (e.g. public relevance, diversity, etc.). A study of content curation processes behind AppleNews, which employs both human moderation (in the Top Stories) and algorithmic content curation (in Trending Stories), showed that human moderated content featured “more diverse and more equitable source distribution than algorithmically-selected” stories.⁵⁹ Trending Stories, according to the same study, almost exclusively included “soft news” (e.g. stories about celebrities), with Top Stories reserved for “hard news” (e.g. political content).⁶⁰ These practices were found to severely impact source plurality, and news distribution, and therefore content plurality, in two ways: First,

⁵⁶ M. E. Mazzoli and D. Tambini, Prioritisation uncovered: The Discoverability of Public Interest Content Online. Council of Europe (2020), p. 23.

⁵⁷ Napoli 2014.

⁵⁸ T. Gillespie, PJ Boczkowski, KA Foot, Media technologies: Essays on communication, materiality, and society, MIT Press (2014).

⁵⁹ J. Bandy and N. Diakopoulos, Auditing News Curation Systems: A Case Study Examining Algorithmic and Editorial Logic in Apple News, Proceedings of the Fourteenth International AAAI Conference on Web and Social Media (ICWSM 2020), p.43.

⁶⁰ Ibid.

aggregators created a so-called “market-expansion effect” because they provided individuals exposure to news outlets with lower popularity or brand awareness. Second, the deployment of aggregators incentivised some users to limit or stop direct use of news outlets, resulting in the so-called “substitution effect”. Since user reactions are usually based on first impressions, clickbait is used in the news feed to attract attention and engage users, thereby facilitating advertising that generates profit. This approach further challenges media sustainability, and consequently independence and pluralism, adding to the overall pressure and financial constraints faced by legacy media because of internet intermediaries’ concentrated advertising and data exploitation models.

There is a growing imbalance between the outreach and communications impact of legacy media and online platforms, and content creation vs. content curation. Given that the traditional subscription-based business model is in decline, legacy media are struggling for viability. An ever-growing number of people get their news exclusively from other sources, where articles are more likely to be available “for free”. The willingness to pay for quality news has decreased, while usage of “free” news aggregator sites and social media platforms increased. As a result, many online news outlets have no choice but to search out new revenue streams. They are coerced into adopting many of the practices used by the large platforms, which exert significant power over the logic of the digital adtech industry (e.g., by employing targeted advertising, publishing sponsored content, or collecting and selling user data). This trend has profoundly negative implications for media freedom globally: It creates an environment in which legacy media organisations must compete with social media companies and intermediaries for the same revenue sources, while also being subject to intermediaries’ recommender systems and content curation policies. The situation contributes to an erosion of trust in media, decreased responsibility for creations and dissemination of disinformation and other problematic content.

Some legacy media organisations also employ algorithm-driven tools themselves, with content personalisation and optimisation playing integral roles in media production processes. There remain stark differences between “news logic of personalisation”⁶¹ and “platform logic of personalisation”, with news media subject to a systemic lack of technological and financial resources, devaluation of traditional editorial and professional ethics, the prevalence of newly emerging private interests, and economic incentives. Algorithmic content curation models and recommender strategies developed and deployed by independent legacy media outlets, and especially public service broadcasters, could offer alternative models for better ensuring audiences’ exposure to diversity, potentially even offering individuals with models for “diversity by design”.⁶²

Evidence shows that journalists value “editorial logic”, such as “transparency, diversity, editorial autonomy, broad information offer, personal relevance, usability, and surprise” over the business-driven algorithmic logic of recommender systems.⁶³

Algorithm-driven content curation and recommendation processes and practices pose threats to media plurality and diversity and raise concerns regarding the full enjoyment of the right to freedom of expression. The major contributing factors constituting these threats are:

- Financial instability of, and fiscal pressure on, legacy media: Online platforms have gained enormous economic power, primarily through advertising revenue, and they use this leverage to dictate conditions

⁶¹ B. Bodó, Selling News to Audiences – A Qualitative Inquiry into the Emerging Logics of Algorithmic News Personalization in European Quality News Media, *Digital Journalism* (2019), p.17-18.

⁶² See more about this content in N. Helberger, *Diversity by design—Diversity of content in the digital age*, Government of Canada (2020), p.8; Natali Helberger, Kari Karppinen & Lucia D'Acunto (2018) *Exposure diversity as a design principle for recommender systems*, *Information, Communication & Society*, 21:2, 191-207, DOI: 10.1080/1369118X.2016.1271900.

⁶³ This study involved newsrooms from the Netherlands and Switzerland; M. Bastian, N. Helberger & M. Makhortykh, *Safeguarding the Journalistic DNA: Attitudes towards the Role of Professional Values in Algorithmic News Recommender Designs*, *Digital Journalism* (2021), p.21.

for the curation of all online content, including editorial media and news content. This power imbalance includes an imbalance of “opinion power”⁶⁴ and the power to “influence processes of individual and public opinion formation”, which in turn enables “these platforms [to] change the very structure and balance of the media market, and thereby directly and permanently impact the pluralistic public sphere.”⁶⁵

- Legacy media, compelled to adopt similar business models and social media “logic”, miss an opportunity to change the “rules of the new communication orders”,⁶⁶ and to contribute to a more diversified media landscape. Yet there are alternative algorithmic-curation models centring on public interest content and professional journalism practices—typically instated by legacy and public service media organisations—and they do offer alternative models to mitigate the potential problems caused by a lack of prioritisation of public interest content.⁶⁷
- Improving algorithmic content curation with a goal of increasing diversity of media sources poses several challenges:
 - Even if internet intermediaries and social media companies “train and game” algorithms “for good”—to expose heterogeneous audiences to heterogeneous content—these practices lack meaningful transparency, and individuals have no agency regarding the design and logic that govern these systems. This presents a significant and systemic risk to the enjoyment of freedom of expression.
 - While personalised content and optimisation processes could contribute to meeting diverse individual, group and societal needs—and generate potential for diversity—such potential should be driven by public policy objectives and corresponding interventions.

⁶⁴ N. Helberger, The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power. *Digital Journalism*, 8(6), 842-854 (2020)

⁶⁵ Ibid., p.846.

⁶⁶ For an in-depth discussion about this problem, see: N. Helberger, The Political Power of Platforms: How Current Attempts to Regulate Misinformation Amplify Opinion Power. *Digital Journalism*, 8(6), 842-854 (2020)

⁶⁷ M.E.Mazzoli and D.Tambini, Prioritisation uncovered: The Discoverability of Public Interest Content Online. Council of Europe (2020).

- There is little to no information on how content produced by and for marginalised communities circulates online, and how recommenders treat such content. Studies suggest⁶⁸ that certain content and speech is treated differently, giving rise to concerns that content is not equally accessible, and that safeguards to prevent discriminatory algorithmic outcomes and ensure fair and equal public participation and deliberation have not been developed or implemented.
- Consequences of algorithm-driven content curation can compound human rights abuses and rule of law violations when amplified in certain national contexts, specifically in conjunction with systemic (state-led and private) media capture and monopolised control of public dialogue. Under these circumstances, additional layers of algorithm-driven restrictions to and for media pluralism and diversity intensify the aggregate sum of individual loss of the right to freedom of expression.

68 See, for example: A. Chinmayi, Facebook's Faces, Forthcoming Harvard Law Review Forum Volume 135 (2021) and K. Klonick, The New Governors: The People, Rules, and Processes Governing Online Speech, The Harvard Law Review (2018); C. O'Neil, Facebook's VIP "Whitelist" Reveals Two Big Problems, Bloomberg Opinion (2021), Retrieved from: <https://www.bloomberg.com/opinion/articles/2021-09-15/facebook-s-xcheck-vip-whitelist-reveals-two-big-problems>

Media Pluralism



3. Human rights-centred recommendations on the use of AI in content curation

States are primarily the guarantor of media pluralism under the international human rights protection framework. They are to act as the ultimate guarantors for the enjoyment of human rights, including responsibility for an enabling environment for the rights to freedom of expression and freedom of the media. The following recommendations for OSCE participating States, generated during the workshop, focus on: strengthening a pluralistic media landscape and plurality of voices (3.1); fostering an enabling environment for the diversity of media content and individual exposure to diverse media (3.2); and enabling individual agency and control (3.3).

3.1 Recommendations on strengthening a pluralistic media landscape and the plurality of voices

This part of the report seeks to offer participating States a normative agenda that fosters a pluralistic media environment and “the coexistence of diverse and competing interests—that is a basis for a democratic equilibrium.”⁶⁹ This agenda is constrained by the shrinking opportunities for a democratic-driven media space, the unbalanced digital platform dominance and excessive market concentration. Participating States should ensure conditions for media innovation, independence and sustainability, especially public interest-driven media, and enforce models of content creation, curation and distribution that foster these conditions.

- **States should ensure, through regulatory initiatives, a level playing field for all media actors, by removing obstacles for the provision of fair and effective market conditions.** The resulting market conditions should enable all media to access and use new technologies and to develop alternative business models—including alternative models of algorithmic content curation that foster a diversified media landscape and the proliferation of public interest content.

⁶⁹ A. Roksa-Zubcevic et al, *Media Regulatory Authorities and media pluralism*, Regional Publication. Council of Europe (2021), p.12-14.

- **States should analyse how existing and future media pluralism-related policy addresses the issue of public interest content**, especially in light of the significance of online platforms in distributing public health information during the COVID-19 pandemic.
- **Public-private partnerships between states and social media companies and other intermediaries should be rigorously transparent** and subject to citizen oversight and public scrutiny. This should include the regulatory framework of the media pluralism landscape.
- **States should employ policy and legislation to prevent the unbalanced and monopolised market power that currently exists**, especially with regards to internet intermediaries and state-controlled content distribution. Any and all state intervention must ensure a pro-democratic regime that is genuinely independent and offers structural solutions to bolster plurality.
- **States should promote plurality and technological and media innovation** by funding holistic independent research that helps media actors, public oversight institutions and academia understand the current distribution of power—especially regarding the effects of recommender systems, analysis of recommender logic, and the resulting impact on media pluralism and diversity.

3.2. Recommendations on fostering an enabling environment for diversity of media content and individual exposure to pluralistic information

This part of the report addresses whether and how internet intermediaries should ensure individuals' equal access to, and participation in, public spaces, by examining diversity as a normative concept.

- **States regulatory and policy interventions should preserve and foster the internet as a space for democratic participation and representation.** Any state regulation of the digital space should

have a clearly defined scope that is necessary for, and proportionate to, a transparent objective, in full compliance with the international human rights framework.

- **States should engage in and support cross-sectoral dialogue to gather the most current and relevant data on the impacts of algorithmic content curation**, such as polarisation, informational gaps, etc. Independent oversight and transparency of diversity monitoring requires a **multidisciplinary approach, led by academic institutions or civil society organisations, with support from the state**. Intersectional diversity monitoring should be used to identify content and audiences that are at risk of, or have historically faced, exclusion from public participation and/or representation.
- **States should adopt an inclusive approach and ensure multi-stakeholder participation and ownership of algorithmic content curation**. Democracies are not self-perpetuating systems. For democracies to thrive, citizens must have the ability to make informed decisions. Through the provision of open dialogue and cross-sectoral collaboration with internet intermediaries, states— together with civil society organisations, marginalised communities, media organisations, journalists and their representatives—could foster sustainable, cross-sectoral cooperation, including between state and non-state actors. Moreover, states should push for diversity in the teams of developers who create algorithmic content curation systems, so that diverse interests and perspectives are represented in the design and implementation of the algorithms.
- **States should provide support and resources to existing independent media regulatory bodies** that use a process of co-creation and inclusion of all national media actors and experts to support an economic, legal and political environment in which diversity is cultivated as a core democratic objective.
- **States should develop an evidence- and research-based legislative framework to ensure accountability of internet intermediaries, including by mandating human rights due diligence**. Human rights impact assessments should be part of any risk mitigation strategy or any external audits to ensure public oversight.

- **States should strengthen independent media regulatory bodies, and other competent institutions, and involve them in public oversight and research.** For instance, these bodies should be involved in human rights impact assessments to address risks that internet intermediaries pose to diversity and plurality, including risks posed to marginalised communities. Such assessments should be accompanied by accountability mechanisms, and there should be transparent disclosure and publication of assessments, audits, and the like.
- **States should increase public funding for independent, quality journalism and/or provide financial resources to independent stakeholders** with relevant expertise and a proven human rights record. These independent actors can offer alternatives to the existing revenue-oriented and data-driven business models, thus fostering decentralised technological algorithmic curation systems that promote public values, such as media diversity, inclusivity and tolerance.
- **States should ensure that any potential intervention in this field does not limit the positive functionality of personalisation or of media independence,** while at the same time providing support and intervention to ensure that content diversity and public interest content is the design-focus. Personalisation can be valuable to individuals when it is used to refine searches and speed up the retrieval of information.
- **States should establish and safeguard adequate data access frameworks that enable vetted researchers, civil society organisations and other independent stakeholders, such as the media, to access data held by internet intermediaries.** At the same time, abuse of such a framework needs to be prevented through the use of ethics guidelines or the creation of an independent authority with an overseer function.

3.3. Recommendation on enabling individual agency and control

The empowerment of individuals should be built-in to algorithmic design, and similarly, public interest and human rights-centred design should be at the forefront of algorithmic deployment.

- **States should support self- and co-regulatory initiatives, and create conditions that codify individual control over what is seen online.** This could be achieved by legally mandating options such as opt-in by default for content recommender systems, and easy identification and choice for defining editorial and non-editorial personalisation.
- **States should mandate transparency and explainability of pre-selected news personalisation and data processing.** This should include transparency of the criteria, principles and types of arrangements driving content prioritisation decisions (to foster public trust, and to allow the public to understand whether commercial or public interest objectives are considered). Participating States should also require intermediaries to have due process measures in place. For example, when intermediaries make restrictions to news feeds they should inform the individual affected about their respective policies and provide effective redress mechanisms. Similarly, **states should support self- and co-regulatory initiatives that ensure intermediary transparency** on the processes driving content prioritisation decisions.
- **States should establish sustainable media and digital literacy programs for all societal groups.** Individuals are often not aware of, and/or do not realise, the implications of algorithmic content curation for their enjoyment of human rights and fundamental freedoms.
- **States should put a special focus on fostering the right to access, seek and impart opinions and ideas of all kinds among all age groups. In particular, states should empower the process of individual opinion formation,** including for young people, who are regularly deprived of proper access to legacy media content.

4. Conclusion

Content curation has, at least from the context of media pluralism and media diversity, largely been pushed to the margins of broader content governance discussions. This omission, compounded by a lack of understanding of the importance of media and informational diversity for heterogeneous audiences, has implications for harm that are equally as concerning as those risks stemming from illegal content and dis/misinformation. All algorithm-driven processes of content curation and moderation are intrinsically connected and must be addressed as such.⁷⁰ These processes are particularly important as algorithms determine what individuals see, what information is prioritised and what content is excluded. Online gatekeepers increasingly rely on recommender systems that systematically analyse patterns of user behaviour and create profiles to determine what information is more likely to engage a given user. In other words, gatekeepers harvest data to determine what personalised content to offer to individual users, in order to spur their engagement and generate more data about them—even if the decisions made are at odds with democratic discourse, the diversity of information, media pluralism and the right to privacy. For this reason, and as clearly articulated in the recommendations, inter-disciplinary research, and transparency of intermediary policies and content curation practices, are crucial preconditions for centring media pluralism and diversity in algorithm design and deployment.

This outcome report highlights the problematic nature of personalised content recommender systems used by internet intermediaries and social media platforms in particular. It outlines the concerning implications these systems have for societal cohesion, diversity, the quality of information within the public discourse and privacy. At the individual level, the online experience is strategically coloured by decisions made for profit, implemented through algorithms without the awareness of affected individuals or the scrutiny of public authorities, and based on intrusive data collection and analysis that are designed to circumvent privacy and data protection laws—with largely negative effects on the diversity of information, media pluralism and the right to privacy.

⁷⁰ In simple words, over-removal of “legitimate” content is in fact also a risk for media diversity.

There exists ample evidence that online platforms' opinion power has the ability to steer and amplify certain public narratives and types of discourse over others. For countries with fragile or oppressive political systems, this opinion power, coupled with algorithmic amplification, can have disastrous consequences for individual enjoyment of human rights. By ranking and differentiating content and recommendation outcomes, internet intermediaries are reconfiguring the public debate in a way that empowers those already in privileged positions. The price for this is a reduction in diversity in general, and particularly disadvantageous for historically marginalised groups, who continue to be pushed to the margins of public discussions in a process that recreates and bolsters inequality and injustice. While algorithmic content curation has the power to limit participation, create division and limit the spread of information, media diversity fosters social cohesion, tolerance and distribution of communication power. It is up to states, primarily, and also non-state actors, most notably intermediaries and media organisations, to ensure that media plurality, equal access and the full enjoyment of human rights are the basis for rules affecting the online information space, as these are building blocks for truly democratic digital societies.

AI in Content Curation and Surveillance-Based Advertising

This part focuses on the use of AI in content curation focusing on the nexus between surveillance capitalism and targeted advertising, and the resulting impact on freedom of opinion and expression. It highlights shortcomings of AI-based content curation and targeted advertising and provides human-rights-centered recommendations for OSCE participating States to address the negative impact that AI tools in content curation have on the right to freedom of opinion and expression.

1. Defining the scope of the impact of surveillance-based business models in their use for content curation

1.1 Impact of automated decision making on the right to freedom of opinion

The international human rights framework distinguishes between the internal and external dimension of the right to freedom of opinion. While the external dimension of this right can be subject to legitimate, proportionate, and non-discriminatory restrictions that are necessary in a democratic society, the internal dimension of the freedom of opinion, so-called *forum internum*, is absolute and non-derogable.⁷¹ Article 19 of the Universal Declaration of Human Rights as well as the International Covenant on Civil and Political Rights protect this absolute right from any restriction or interference. In the words of the UN Special Rapporteur on Freedom of Expression and Opinion, “any involuntary disclosure of opinions is prohibited and mental autonomy is affirmed.”⁷²

The data-harvesting business models of large online platforms enable the

⁷¹ Office and the High Commissioner for Human Rights, CCPR General Comment No. 22: Article 188 (Freedom of Thought, Conscience and Religion), available at <https://www.refworld.org/docid/453883fb22.html>, 1993.

⁷² Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Irene Khan, Disinformation and freedom of opinion and expression, available at: <https://undocs.org/A/HRC/47/25, 2021>.

advertising industry to develop or rely on data-driven targeting strategies. Through this approach, companies identify and exploit people's or communities' behavioural patterns and characteristics. The umbrella term that covers these manipulative techniques is "surveillance-based advertising", understood as a blanket term for digital advertising that is targeted to individuals or groups, usually through tracking and profiling based on personal data. The context of where a specific ad is placed can be random, because, as it is targeted at individuals, it can follow them around in different contexts.⁷³ In most cases, surveillance-based advertising is part of an automated process, by which each individual ad is chosen and placed in a matter of milliseconds. This means that neither the ad publisher (e.g. the owner of a website or app) nor the advertiser (e.g. the owner of the brand that is promoted) chooses which ads to show to whom, or where to display them. This is automatically decided by technological systems that are often controlled by third party intermediaries (so-called "adtech" companies).⁷⁴

Surveillance-based advertisement has significantly contributed to the exploitation of people's particular characteristics to increase the persuasiveness of a message, thereby unjustifiably interfering with their absolute freedom to form an opinion and to enjoy independent thought processes. People who are using platforms' services are being manipulated to think or to make decisions they would have otherwise perhaps never made. Surveillance-based advertisement exploits individuals' vulnerabilities even if it does not directly identify those vulnerabilities. Through the use of so-called "lookalike audiences", advertisers can duplicate people's groups with certain characteristics in order to reach new individuals that share the same characteristics. Automated tools and the dominance of a few online platforms has enabled greater manipulation as every single individual using their service can be targeted all the time and at any time.

⁷³ Norwegian Consumer Council, Time to ban surveillance-based advertising: The case against commercial surveillance online, available at: <https://www.forbrukerradet.no/wp-content/uploads/2021/06/20210622-final-report-time-to-ban-surveillance-based-advertising.pdf>, 2021.

⁷⁴ Norwegian Consumer Council, Out of control: How consumers are exploited by the online advertising industry, available at: <https://fil.forbrukerradet.no/wp-content/uploads/2020/01/2020-01-14-out-of-control-final-version.pdf>, 2020.

There is increasing demand for a ban on practices that adversely impact people's absolute right to freedom of opinion and freedom of thought, in particular as individuals' thoughts and opinions are being widely influenced without their knowledge or consent. This phenomenon specifically includes targeted behavioural tracking and individual cross-site/cross-device tracking. Such constant invasive corporate surveillance poses a risk of systematic manipulation of individuals, beyond traditional forms of advertising influence. Surveillance-based advertising targets individuals in opaque ways⁷⁵ and may exploit vulnerabilities, opening new possibilities for manipulation. In particular, when combined with revenue-maximising algorithms for content curation, surveillance-based advertising may impact how individuals speak out and behave, effecting diversity of information, views and opinions.

Despite online platforms' claims that there is no turning back from surveillance-based advertising, the internet was not built on a "creepy ad" business model. In fact, quite the opposite. States must avoid directly or indirectly protecting business models that stand on surveillance-based advertisement and violate international human rights law. Ending abusive models also means opening the door to human rights compliant alternatives, including innovative forms of contextual advertising that rely on minimum personalisation and no individual targeting.⁷⁶ This will also enable new players to enter the digital market.

Surveillance-based advertisement has far-reaching impacts on people's personal interactions, choices and participation in democratic debates. Measures intended to increase transparency can help to better understand the scale of the issues, but these are not enough to prevent and mitigate the ongoing human rights abuses. The individual and societal harms created by intrusive targeting and personalisation require a systematic response. From privacy intrusions to content curation, invasive tracking harms the right to freedom of opinion in tangible ways. It is a positive obligation of states to protect this absolute right from such interferences by creating an adequate regulatory framework establishing and enforcing strong human rights safeguards.

⁷⁵ Civil society efforts to gain more transparency have confronted obstacles: <https://algorithmwatch.org/en/defend-public-interest-research-on-platforms/>.

⁷⁶ Natasha Lomas, Data from Dutch public broadcaster shows the value of ditching creepy ads, available at: <https://techcrunch.com/2020/07/24/data-from-dutch-public-broadcaster-shows-the-value-of-ditching-creepy-ads/?guccounter=1>, 2020.

1.2 Guiding note on online targeting

While many intermediaries target their users (as well as non-users) via behavioural profiling and cross-site tracking, online gatekeepers with unprecedented access to large amounts of users' data are leading the adtech industry. In practice, surveillance-based advertising starts with an ad publisher who operates a website or a mobile app that delivers a service or content. They provide a space for placing ads on their platforms and/or access to data about their users. Their trading partners are marketers, companies that are eager to sell their products to the most valuable customers. But third-party vendors and online ad-exchanges stand in between these actors. They operate in the shadows, do not have any direct relationship with users, decide which ads will be placed on which sites, and receive a part of the transaction. This complicated advertising network collects, analyses and merges extensive amounts of personal data without people's knowledge. Neither publishers nor marketers are able to fully or even partly control this process.

Major internet intermediaries assert market dominance across the advertising ecosystem by holding all three roles simultaneously – acting as ad publishers, marketers and third-party vendors. Their dominance is further reinforced by their virtually limitless access to data, including data from their own services and third party data. This creates an enormous power imbalance that fuels unfair competition in the digital market and poses a risk of systemic human rights abuse.

The main focus of this part of the report is internet intermediaries whose business models heavily rely on online targeting. While many intermediaries target their users via behavioural data and cross-site tracking, which amount to human rights intrusive practices per se, these online gatekeepers with unprecedented access to large amounts of users' data are also leaders of the adtech industry. For instance, large social media platforms such as Facebook have developed very granular systems for their advertising interface thanks to which they control major advertising revenue globally. The digital rights organisation Panoptikon mapped and described in depth the adtech ecosystem developed by the gatekeeper Facebook, as well as its impact on human rights. Panoptikon points out that Facebook is not only a passive intermediary between

advertiser and users.⁷⁷ It enables advertisers to select criteria that are then interpreted by Facebook's algorithm in order to achieve the advertiser's desired objectives.

This report uses Facebook as a practical example to demonstrate how surveillance-based advertisement works in practice. First, advertisers are able to select their target audience based on targeting criteria that are determined by the intermediary. There are a number of criteria that advertisers can use. Among others,⁷⁸ advertisers can choose to target a custom audience or so-called lookalikes. Both criteria were introduced by Facebook in recent years and are often described as follows:

- **Criterion of custom audience** is based on an advertiser's own information that they hold about their users and can upload to the intermediary. Consequently, the intermediary's algorithm matches this information with its own data about the users – without revealing the users' profiles to advertisers.
- **Criterion of lookalike audience** enables advertisers to target a group of users who are similar to the originally desired one. In practice, the intermediary predicts what audience shares the characteristics with the original targeted group – the so called “seed audience”.⁷⁹ Lookalikes are identified by the intermediary's matching algorithm.

Online platforms are able to target individuals with high precision because they possess data and knowledge about individual users and non-users alike. Big data analysis allows them to predict individuals' behaviour, using data that is directly provided by users or obtained by observing online activity and behavioural patterns of users and others. Highly sensitive algorithms create profiles based on behavioural data—habits, preferences, dislikes and interactions with users. These profiles may even include conclusions drawn from the times that users are most active online.

⁷⁷ Panoptikon Foundation, Who (really) targets you? Facebook in Polish election campaigns, 2020.

⁷⁸ Panoptikon Foundation, Who (really) targets you? Facebook in Polish election campaigns, 2020.

⁷⁹ Panoptikon Foundation, Who (really) targets you? Facebook in Polish election campaigns, 2020. See also, Norwegian Council for Consumer Protection, Out of control: How consumers are exploited by the online advertising industry, 2020.

Both the creation and subsequent use of profiles are privacy invasive and involve assumptions, and they can lead to discrimination. Algorithms are also capable of inferring further information about individuals that the targeted individual does not intend to reveal. The rationale behind this technique is the notion that the more companies know about their users, the greater the likelihood that they can successfully predict and potentially manipulate them. This information is then used to deliver specific content and advertising “at the right time and in the right context,” to incite users to buy certain products or services, or to watch certain videos.⁸⁰

80 Vladan Joler, *The Human Fabric of the Facebook Pyramid*, SHARE Lab Foundation, 2017.

Surveillance



2. Human rights-centred recommendations on regulation of surveillance-based advertisement

2.1 Recommendations to strengthen users' empowerment and personal agency in online ecosystem

- **States should take a people- and user-centred approach to strengthening user empowerment, individuals' agency and control over their data.** There is a significant risk created by our inability to know if we have been profiled or identified, how we have been profiled or identified, and by what algorithm. Yet, transparency alone is not enough for people to control the use of their information. Transparency should be coupled with robust, actionable rights to reject such practices.
- **States should ensure that strengthening users' empowerment and personal agency is not mutually exclusive with constructing complementary systems of external oversight and inquiry.** It is important to prioritise user agency and control in the current socio-political environment.
- **States should invest in research** to develop an empirical foundation for identifying and understanding the effects of surveillance-based advertising on user autonomy and agency. Without further empirical studies, there may be an oversimplification of users' experiences based on scarce data points and research that is focused on particular online communities.
- **States should promote a regulatory framework to improve information distributed to users,** to allow users to exercise free choice in the advertisements they view and to which they respond. The framework should also ensure that users are more aware of the data collected about them and how it is used (including the reasons why a specific user is targeted for a specific advertisement).

- **States should clarify where existing media and content regulation applies to virtual content.** Where gaps are identified, states should review and develop policies and recommendations to moderate online content in the context of the inaccessibility (or “black boxing”) of online ecosystems and platforms.
- **States should promote business practices that provide alternatives to current surveillance-based advertisement.** Current business practices of internet intermediaries create a problematic concentration of power that negatively impacts users’ personal autonomy and agency.
- **States should ensure private actors act in accordance with the UN Guiding Principles on Business and Human Rights** so that corporate values and governance structures do not prioritise profit maximisation at the expense of human rights and democratic values. The public is increasingly demanding that businesses do not operate in a commercial vacuum but reflect democratic values and priorities.
- **States should encourage the private sector to pursue non-legal avenues to promote greater transparency and accountability.** Private initiatives on codes of ethics play a crucial role in corporate social responsibility. In isolation, however, such self-regulatory approaches alone cannot provide effective protections against the potential for surveillance-based advertising to infringe on the absolute right to freedom of opinion.

Recommendations on outreach and raising awareness about surveillance-based advertisements among the general public

- **States should promote awareness and digital literacy** so that individuals know how to manage their own media consumption and use of internet intermediaries. Users should have a meaningful understanding of why they are shown targeted content as well as how their personal data is being processed and accessed. It is important

for users to understand not only the amount of their personal data that is being processed, but also the type of information that can be accessed, who can access it, and the way certain information can be linked to protected characteristics. Increased digital literacy is important to empower users and strengthen resilience to an ever-adapting industry.

- **States should recognise how surveillance-based advertisement can impact rights to equality and non-discrimination in combination with the right to freedom of opinion and expression.** Surveillance-based advertising creates different and potentially discriminating experiences, both within and between groups of people sharing certain characteristics. This can be exacerbated because certain groups lack digital literacy and that can compound negative experiences with surveillance-based models.
- **States should encourage private actors to consider the concept of “social licence”,** which seeks to ensure that private and public service providers act responsibly and ethically in the best interests of the community.
- **States should invest in research** to develop a strong empirical foundation that can ensure outreach and awareness initiatives addressing practical issues of how the public responds to online manipulation and targeted advertising strategies.

Recommendations for legally mandated meaningful transparency: different layers of transparency

- **States should ensure meaningful transparency of surveillance-based advertising.** Personalisation of surveillance-based advertising means that different individuals see different ads based on a number of factors, including time of day, context, demographics, personal characteristics and behavioural patterns. Yet algorithmic systems that are being fed with users' data are profoundly opaque, often described as being “black-boxed”. Hence, the decisions behind surveillance-based advertising are near to impossible for users (or regulators) to understand. As a consequence, users lack any meaningful comprehension of why they are being shown a particular

ad at a particular point in time, and how their personal data is shared and used in the process.

- **Designated oversight bodies with expertise in the areas of equality and non-discrimination should be empowered to monitor and address the unequal or discriminatory effects that surveillance-based advertising has on marginalised groups.** States should consider various approaches to accountability for harmful surveillance-based advertising. States should consider the merits of self- and co-regulation models, corporate accountability and governance, litigation mechanisms or alternative e-courts in creating responsibility for adverse consequences.
- **States should ensure that equality bodies are empowered to undertake strategic litigation** to challenge discriminatory outcomes of automated measures.
- **States should collaborate with academia, civil society and independent stakeholders** to refocus transparency efforts on achieving greater access to large-scale disaggregated data that can enable research and understanding of data-driven profiling and advertising. Transparency is required for states and the public to know how surveillance-based advertising is deployed. This will enable meaningful research and allow for challenges to problematic processes. Data access for meaningful public interest research should be based on a legal framework.

Recommendations tackling the interplay between individual versus group privacy

- While international human rights law defines individual rights, online profiling has collective aspects and impacts. Digital profiles are based on inferences and assumptions about a complex web of data and networks. Algorithmic profiling can correlate characteristics and connections to profile individuals from marginalised groups. **States should therefore ensure internet intermediaries respect the right to freedom of opinion and are aware of its important intersection with the rights to freedom of association and expression.**

- **States should consider the constraints of existing legal mechanisms for enforcing collective rights in the context of surveillance-based advertisement.** Current legal systems reflect individual rights, and the only extension to group concerns exists when an individual belongs to a certain group. In these cases, even when an individual makes an informed choice to opt-out of sharing their personal data, they may still be profiled as a part of a wider group that is targeted or categorised by AI systems. States should ensure protection and regulation of the use of personal data, including metadata or demographically identifiable data, which is considered extremely relevant and valuable when it comes to advertising methods.

2.2 Recommendations to develop regulatory and co-regulatory solutions that can effectively address negative impact on human rights stemming from surveillance-based advertising

Recommendations to safeguard the absolute right to freedom of opinion

- **In line with international human rights standards, states should respect and promote the absolute right to freedom of opinion**—which includes the rights to keep one's thoughts and opinions private, to not have one's thoughts and opinions manipulated, and to not be penalised for one's thoughts and opinions.
- **States should emphasise that everyone enjoys the right to hold opinions without interference; the right to seek, receive and impart information and ideas through any media, regardless of physical frontiers; and the right to not be subjected to unlawful or arbitrary interference with their privacy.** States should ensure that individuals can form their opinion while being protected from manipulation by opaque profiling methods that determine when a user is most susceptible to behavioural influence in order to exploit the user's vulnerabilities. Undue influence can stem from practices such as: non-transparent or non-verifiable targeted advertising at scale; tracking and behaviour observation techniques or obfuscating design features

(“dark patterns”); or the use of power imbalances to influence thoughts (speed, scale, inaccessibility, “black-boxiness” and systematic non-transparent influence). Such tracking and targeting techniques can lead to self-censorship and conforming effects, which might be more prevalent among certain segments of the population. States should set clear policies and criteria for the line between legitimate influence and illegitimate manipulation based on algorithmic technologies, for which states should consider moratoriums or bans.

- **States should consider the legal recourse for users** to address penalisation imposed by internet intermediaries through flaws in the system, or through user classification, which impacts an individual’s online experience regardless of the accuracy of the classification.
- **States should legally ensure anonymity and encryption**, and this should include ensuring that opinions are not disclosed involuntarily.
- **States should invest in digital and media literacy and education campaigns to inform the public about how targeted advertising based on profiling and surveillance methods impacts individuals’ online experience, and how surveillance advertising threatens freedom of opinion.** Constant surveillance is not in accordance with human rights, and it risks creating chilling effects on freedom of opinion and expression. States should ensure that individuals have sufficient tools and information diversity to form their opinions freely and to enjoy the positive aspects of freedom of thought and opinion.
- **States should address surveillance-based advertising in the sociotechnical context of content moderation and content curation**, and should identify ways to address the centralisation of power, including due to the bundling of several services.

Recommendations for meaningful transparency and for regulatory measures of online targeting

- **States should develop a human rights policy with emphasis on salient human rights issues**—such as freedom of expression, freedom of the media, privacy and freedom from discrimination.

States are the duty-bearers under international human rights law and hold a positive obligation to protect human rights from interference by others, including by private actors or individuals. States should therefore commit to adhering to international human rights law, and should ensure that national laws and policies regulating internet intermediaries and the advertising industry are fully compliant with the international human rights framework.

- **States should ensure private actors act in accordance with due process and the standards of legality, legitimacy and acceptance of oversight by an independent and impartial judicial body, in line with the UN Guiding Principles on Business and Human Rights.** States should establish a regulatory framework for companies to demonstrate that they have rigorously implemented their responsibilities under the Guiding Principles.
- **States should effectively enforce existing data protection and privacy laws.** In this context, states should provide for principles such as data minimisation and purpose limitation. States should also effectively enforce competition and antitrust laws, as well as other regulations aimed at strengthening human autonomy.
- **States should condition corporate surveillance—including targeted advertising that uses tracking and profiling—on human rights due diligence** and a track record of compliance with the UN Guiding Principles on Business and Human Rights.
- **States should oblige internet intermediaries to provide documentation about AI-based tracking and profiling methods that they deploy for advertising purposes.** States should require internet intermediaries to provide explanations regarding the models used, what data is collected, and for what purpose—as well as performance metrics and testing results for the models used. States should mandate that internet intermediaries must properly explain how their advertising and business models work, how algorithmic decision-making is involved, and how such automated systems make decisions affecting the user. Any disclosure should be made in a way that is understandable and accessible to users. Information on the collection of personal protective characteristics, or proxies of these,

should be included. Additional information should be shared, in a privacy-friendly manner, with researchers and regulators.

- **For any data-harvesting and advertising-based business models, states should require ex ante human rights impact assessments** that are part of a clear regulatory framework, and are transparent, independent and inclusive (involving meaningful consultation with potentially affected groups and other stakeholders). The process should include oversight by a regulatory agency or independent stakeholders with relevant expertise, to ensure the mitigation of adverse impacts of advertising models on prevention of discrimination and preservation of freedom of opinion and expression.
- **States should adopt new constraints, or enforce existing ones, that limit what types of data can be collected, and how it can be used, and what types of data may be disclosed to advertisers, data brokers or third parties.**
- **States should clearly define the way advertising methods cause “harm”** (individually as well as collectively/to democratic processes), based on the precautionary principle, so they can identify a threshold for banning harmful surveillance-based advertising practices. Such bans should include, for instance, forbidding weaponising of sophisticated techniques for influence based on psychological models that assume psychological vulnerabilities and manipulability. For data harvesting for targeted advertising that is within the threshold, states should ensure strict transparency methods – for instance regarding product placement – and human rights due diligence, which puts the best interest of the individual at centre.
- **States should ban indiscriminate mass collection and analysis of user data for targeted advertising that harms users individually or collectively, or interferes with their right to freedom of thought and opinion.** This includes, for example, targeted advertising based on pervasive tracking of users' vulnerabilities or categories of protected characteristics, such as ethnicity, gender, religious belief or sexual orientation. Bans and restrictions of surveillance-based advertising could follow the model of bans on deceptive and subliminal advertising or restrictions on advertising alcohol, tobacco, gambling

or environmentally hazardous materials. Special protection should be considered for vulnerable/susceptible groups, such as children and young people.

- **States should ensure that personalised advertising using scraping of personal data operates on the basis of informed consent and on an opt-in basis.** States should ensure users are able to make choices about which data is collected for which purpose and how they want to engage in online debate and be targeted with advertising (including seeing personalised advertising and being surveilled for advertising in the first place). For less invasive advertising models, at least an option to opt-out of data collection should be provided, and there should be an alternative means for ensuring users' safety online. Consent needs to be explicit, non-coercive and based on informed choice, complying with data protection laws, while acknowledging that advertising models can impact human rights not only by collecting and analysing personal data, but also by using other information and metadata. Users should have control over which data is collected, retained or inferred, and how it is used for advertising. States should promote privacy by design and by default.
- **States should mandate that internet intermediaries must provide information about their revenue model and ensure a network for transparency.**
- **States should oblige internet intermediaries to notify users when they are subjected to any form of tracking and profiling,** to tell users how such mechanisms operate, and to provide for opt-in or opt-out options in an easy and user-friendly way. States should mandate that intermediaries must disclose whether advertising content is shown based on users' own history, location information, social media activities, demographic characteristics or other information (including proxies and "lookalike audiences" that group users with certain characteristics). Intermediaries should also be required to disclose their targeting parameters and audience categories (based on behaviour as well as content), as well as the guidelines against which audience categories are evaluated, and whether algorithmically generated categories are reviewed by human reviewers before being used.

- **States should ensure that users have access to profiling data that internet intermediaries hold about them**, as well as any inferences made about them (including metadata, such as assigned categories and the list of advertisers attempting to influence them). This data should be made available to users upon request, in a comprehensible and accessible format. Users should be able to rectify and delete their profile.
- By introducing certain bans and mandating transparency about what data is collected, stored and analysed, and what kind of advertising decisions the data is used for, **states should address the opaque surveillance-based advertising that may impact the ability of individuals to use internet intermediaries' services as forums for free expression, access to information and engagement in public life.**
- **States should require regular transparency reporting**, mandating minimum requirements on the data collected, categories used and automation involved – and how these impact content and advertising provided. Intermediaries should also be required to provide mandatory, functional advertising libraries.
- **States should put in place a framework for internet intermediaries to disclose their human rights impact assessments and ensure external independent review.** These should include assessments of how freedom of expression and information risks are associated with their targeted advertising policies and practices, as well as assessments of discrimination risks.
- To ensure independent external audits of advertising models, **states should require internet intermediaries to conduct reporting in line with privacy and data protection and is accessible to all relevant public authorities and independent stakeholders**, including researchers and civil society organisations.
- **States should require that internet intermediaries grant researchers and civil society organisations access to their advertising data**, so they can evaluate advertising practices, and their individual and collective impact, and inform public interest-driven research.

- **States should ensure democratic governance**, and recognise and empower designated oversight bodies that have expertise in the areas of equality and non-discrimination to monitor and address unequal or discriminatory effects of surveillance-based advertising on marginalised groups.
- **States should strengthen the independence of data protection bodies** and provide them with sufficient political support and financial resources and competencies.
- **States should encourage multi-stakeholder coordination**, capacity-building and research into the impact of interface design on user behaviour, as well as issues such as “dark patterns”. States should also promote research into marginalisation and the gendered nature of digital surveillance and advertising effects – and research of negative externalities of business models based on collection of personal information at a planetary scale, which allows micro-targeting of individuals tailored to their specific attributes, traits and preferences. In addition, research should investigate: targeted advertising’s potential to influence behaviour maliciously; the connection between surveillance-based business models and intermediaries’ incentive to prioritise harmful content, enabling discrimination harms enacted through algorithmic decision making; and associated chilling effects.
- **States should refrain from arbitrarily accessing data collected by internet intermediaries.** Data requests should be based on legitimacy, legality and necessity, and proportionality, and should provide for judicial oversight. States should adequately enforce safeguards to prohibit mandatory transfer of data, especially to law enforcement, and take specific measures to protect marginalised and vulnerable groups.
- **States should address the concentration of power**, which includes the harvesting of data as a source of market power and further reinforces the dominance of a few dominant intermediaries to the detriment of potential competitors and news publishers. Measures could include, for example, mandating interoperability, data portability (data ownership) through secure mechanisms and/or decentralising power.

- **States should address the impact that the concentration of the digital advertising market has on legacy media and the availability of information of public interest.** States should invest in strong public service media and independent journalism.
- **States should invest in exploring alternative revenue streams that do not rely on commodifying people's private behaviour and do not influence or shape emerging behaviour.** Examples of such alternatives include contextual advertising or targeting according to simple criteria to which a user opts in, direct links between content providers and advertisers without the intermediary ad sector monetising content edited by others (including the media), and efforts to encourage human rights-friendly innovation.
- **States should consider public service platforms** that serve and are fully accountable to the public, based on democratic governance.

2.3 General principles for preventing states from piggybacking on surveillance-based business models

Data-harvesting business models connected to surveillance-based advertising can be abused by states as well. Public authorities increasingly rely on data mining from private companies, which serve as “reservoirs of consumer data”. Governments are regularly able to access a variety of data provided by the private sector. In recent years, there were a number of cases reported by civil rights groups documenting how public authorities struck an agreement with data brokers in order to gain access to users' personal data. For instance, the Electronic Frontier Foundation (EFF) described the case of the U.S. Immigration and Customs Enforcement buying ALPR data⁸¹ from Vigilant to help locate people the agency intends to deport.⁸² Such informal agreements between governments and private entities pose a serious threat to the protection of human rights. Unnecessary and disproportionate surveillance may undermine security online and hinder access to information and

⁸¹ <https://www.techdirt.com/articles/20190321/09165441842/vigilant-customers-are-lying-about-ices-access-to-plate-records.shtml>.

⁸² <https://www.aclunc.org/blog/documents-reveal-ice-using-driver-location-data-local-police-deportations>.

ideas.⁸³ Surveillance may create a chilling effect on the online expression of individuals, particularly journalists and members of civil society, who may self-censor for fear of being constantly surveilled. Moreover, surveillance exerts a disproportionate impact on the freedom of expression of marginalised groups, including racial, religious, ethnic, gender and sexual minorities, as well as journalists or human rights defenders.⁸⁴ This holds equally true for state surveillance as it does for corporate surveillance.

In particular, advanced development of AI has enabled new possibilities of en masse state surveillance that piggybacks on the architecture of intermediaries' business models. Different forms of content monitoring tools deployed by states to clarify relationships between targeted users, or to assign a meaning or attitude to their social media posts via natural language processing and sentiment analysis, may have serious repercussions for the protection of human rights online. When this process is empowered by machine learning, states can uncover connections and interlinks that are potentially invisible to the human eye. Especially in authoritarian regimes, human rights defenders, political activists and the marginalised may be persecuted for their opinions and views, leading to disproportionate and severe punishments.

This part contains general principles that states should follow, in order to prevent human rights abuse at scale:

1. **Public authorities, and especially law enforcement agencies, should have very limited and specifically targeted access to data, narrowed to specific identifiers or specific categories.**
2. **Data collection by law enforcement should always be based on concrete suspicions.** Law enforcement should only obtain access to specific records and content. No bulk monitoring should be performed, including facial recognition that can enable mass surveillance.
3. **Data collected using special national security powers should not be used for any other government purpose, including law enforcement.** It should be retained for a limited period and deleted once no longer required.

⁸³ A/HRC/23/40, <https://undocs.org/en/A/HRC/23/40>.

⁸⁴ A/HRC/29/32, <https://undocs.org/en/A/HRC/29/32>.

4. **Metadata revealing information, such as who people communicate with, and where and when, can be extremely revealing about individuals' lives, and thus should receive a high level of legal protection.**
5. **Illegal surveillance should be criminalised, with effective remedies.**
Illegally gathered data should be inadmissible as evidence, while whistleblowers should be protected when revealing illegal behaviour.

3. Conclusion

This part of the report highlights the impact of AI-enabled and surveillance-based targeted advertising and data-harvesting business models on content curation, information plurality and the ability of individuals to form, hold and express their opinions, and access information freely.

It analyses the link of targeted advertising to the rise of powerful internet intermediaries, who simultaneously act as gatekeepers to expression and information in the digital marketplace of ideas. It also examines how the value and thus prominence of online content is increasingly made dependent on its contribution to generating advertising profit for intermediaries. Illustrating how individuals' data, characteristics and vulnerabilities are exploited for targeted advertising, the report outlines the impact of profit considerations on content governance and online information spaces. The report explores how the current digital ecosystem may interfere with the absolute right to freedom of opinion and with the right to seek, receive and impart information of all kinds, regardless of physical frontiers.

Moreover, the report highlights the link of data-harvesting business models of internet intermediaries to state surveillance. Surveillance creates a chilling effect on online expression of individuals, and journalists and civil society in particular, with a disproportionate impact on marginalised individuals and groups. This holds true for state as well as corporate surveillance.

The report outlines a set of proactive, preventative and responsive recommendations for OSCE participating States. These human rights-

centred recommendations focus on safeguarding the absolute freedom of opinion, on ensuring meaningful transparency, on providing regulatory measures of online targeting, and on general principles for preventing states from piggybacking on surveillance-based business models. While certain challenges, such as the lack of explainability, transparency and accountability of AI-based systems linked to advertising and content governance, need to be addressed urgently, the report also identifies the need to address the broader surveillance-based ecosystem in order to genuinely protect and promote freedom of opinion and expression in the digital age.