Council of the European Union
General Secretariat

Brussels, 18 March 2022

WK 4035/2022 INIT

LIMITE

COMPET	**PI**
MI	**AUDIO**
JAI	**CONSOM**
TELECOM	**CODEC**
CT	**JUSTCIV**
	TTC

This is a paper intended for a specific community of recipients. Handling and further distribution are under the sole responsibility of community members.

NOTE

From:	European Commission
To:	Delegations
Subject:	Algorithmic Amplification: Input paper for Workshop 2 EU-US Tech and Trade Council Working Group 5

Algorithmic Amplification

Input paper for Workshop 2
EU-US Tech and Trade Council Working Group 5
4 March 2022

This workshop will focus on understanding concerns related to algorithmic amplification, with a special focus on content-sharing algorithms on technology platforms, addressing the merits of these systems as well as their potential negative effects. What voluntary actions have social media platforms and others taken to mitigate any harms? What are the limitations of, and obstacles to, these practices? And what improvements to these practices are contemplated, useful, and/or practical?

1. The effects and impact of algorithmic amplification

Algorithmic amplification is an umbrella term covering different computational processes. For the purposes of this workshop, this concept would refer to any platform's use of an algorithm, model, or other computational process to rank, order, promote, recommend, or similarly alter the delivery or display of information (including any post, page, group, account, channel, or affiliation) provided to a user of the service that achieves a business objective for the company, such as tailoring marketing content to advertiser and user preferences. Algorithmic amplification is only one form of amplification; it is only one part of a complex socio-technical system where users' behaviours influence and are influenced by the propagation of information online. The modification of social media algorithmic practice can be related to more general policies and practices around algorithms and AI.

Content may be harmful when viewed by relatively few people under certain circumstances. It may also be harmful when "disseminated to a large audience [which] can contribute to systemic problems."[1] It is important to acknowledge the range of unknowns around the impacts of algorithmic amplification, including its potential benefits, given the limited access to information about the algorithms and their use and data on how these computational processes impact individuals, communities, and societies.

Research shows that algorithmic amplification can have a snowball effect, as one piece of content encourages others to produce and share similar content, exponentially increasing the numbers exposed. Guillaume Chaslot has described the dangers of this feedback loop in certain contexts: "once a conspiracy video is favored by the A.I., it gives an incentive to content creators to upload additional videos corroborating the conspiracy. In turn, those videos increase the retention statistics of the conspiracy. Next, the conspiracy gets recommended further. Eventually, the large amount [sic] of videos favoring a conspiracy makes it appear more credible."[2]

Individuals seek out information that confirms their established opinions and biases. Recent research demonstrates that personalization algorithms tend to funnel many users toward ideological extremes

[1] Jennifer Cobbe and Jatinder Singh, "Regulating Recommending: Motivations, Considerations, and Principles," *European Journal of Law and Technology*, 10:3 (2019).
[2] Matthew Ingram, *Fake news is part of a bigger problem: automated propaganda.* Columbia Journalism Review, 22 February 2018, https://www.cjr.org/analysis/algorithm-russia-facebook.php

and can increase polarization.[3] By nature of their design, recommendation algorithms run the risk of exacerbating existing tendencies among media consumers and platform users to insulate themselves from exposure to different viewpoints.[4] The dynamics of recommendation algorithms may compromise "the ability for lay publics to ascertain the veracity of claims to truth."[5]

Debate exists over the extent of these effects.[6] While some studies have challenged the view that recommendation algorithms are necessarily radicalizing,[7] others have confirmed the notion of a "radicalization pipeline"[8]or "immersive ideological bubble"[9] in YouTube recommendations, as well as higher rankings given to extreme or fringe content.[10] The creation of self-reinforcing biases and "filter bubbles" are damaging to the normal functioning of public debate, group deliberation, and democratic institutions more generally.[11] While filter bubbles exist outside the online environment, algorithmic targeting and amplification create unique challenges of scope, reach, and precision based on personal data and profiles. Eight of eleven studies in an overview of research examining algorithms and terrorism-related content found that algorithms such as those used in YouTube recommender systems or Facebook's suggested friends amplified extremist content and that algorithmic systems direct harmful content to the vulnerable, including children and adolescents, threatening mental health and impeding child development.[12] Platforms frequently become vehicles for the spread of harmful health-related misinformation and disinformation, with users flocking to divisiveness.[13] Internal research from Facebook showed that a change to the News Feed algorithm in

---

[3] Ivan Dylko et al. "Impact of Customizability Technology on Political Polarization," *Journal of Information Technology & Politics*, 15:1 (2018): 19-33; Jaeho Cho et al. (2020) "Do Search Algorithms Endanger Democracy? An Experimental Investigation of Algorithm Effects on Political Polarization," *Journal of Broadcasting & Electronic Media*, 64:2 (2020): 150-172; Silvia Milano, Mariarosaria Taddeo, Luciano Floridi, "Recommender systems and their ethical challenges," *AI & Society,* 35.4 (2020): 957–967.

[4] Christopher Bail, "Exposure to opposing views on social media can increase political polarization," *Proceedings of the National Academy of Sciences* 115:37 (2018): 9216-9221.

[5] Joan Donovan & Danah Boyd, "Stop the Presses? Moving from Strategic Silence to Strategic Amplification in a Networked Media Ecosystem," *American Behavioral Scientist,* 65:2 (2021): 333-350.

[6] Seth Flaxman, Sharad Goel, and Justin M. Rao, "Filter bubbles, echo chambers, and online news consumption," *Public opinion quarterly,* 80.S1 (2016): 298-320.

[7] Chris Bail, *Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing*, Princeton University Press, 2021; Mark Ledwich & Anna Zaitsev, "Algorithmic Extremism: Examining YouTube's Rabbit Hole of Radicalization," *arXiv Preprint*, 1912.11211(2019).

[8] Manoel Horta Riberio et al., "Auditing Radicalization Pathways on YouTube," *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 2020.

[9] Derek O'Callaghan et al. "Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems." *Social Science Computer Review*, 33:4 (August 2015): 459–78.

[10] Joe Whittaker et al. "Recommender systems and the amplification of extremist content", *Internet Policy Review,* 30 June 2021.

[11] Engin Bozdag, "Bias in algorithmic filtering and personalization," Ethics and Information Technology, 15:3 (2013): 209-227.; Engin Bozdag & Jeroen van den Hoven, "Breaking the filter bubble: democracy and design," *Ethics and Information Technology,* 17.4 (2015): 249-265; Jaron Harambam, Natali Helberger, & Joris van Hoboken, "Democratizing algorithmic news recommenders: how to materialize voice in a technologically saturated media ecosystem," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376.2133 (2018): 20180088; Natali Helberger, Kari Karppinen, & Lucia D'acunto, "Exposure diversity as a design principle for recommender systems," *Information, Communication & Society*, 21.2 (2018): 191-207; Ansgar Koene et al., "Ethics of personalized information filtering," *International Conference on Internet Science* (2015): 123-132; Urbano Reviglio, "Serendipity by Design? How to Turn from Diversity Exposure to Diversity Experience to Face Filter Bubbles in Social Media," *International Conference on Internet Science* (2017): 281-300; Matthew Zook et al., "Ten simple rules for responsible big data research," *PLOS Computational Biology*, 13:3 (2017): e1005399.

[12] *Content-Sharing Algorithms, Processes, and Positive Interventions Working Group Part 1: Content-Sharing Algorithms & Processes,* Global Internet Forum to Counter Terrorism, July 2021: https://gifct.org/wp-content/uploads/2021/07/GIFCT-CAPI1-2021.pdf

[13] Sylvia Chou, Wen-Ying, & Anna Gaysynsky, "A prologue to the special issue: health misinformation on social media," *American Journal of Public Health*, 110.S3 (2020): S270-S272.

2018 towards boosting interactions with friends and family resulted in higher rates of polarization and outrage, amplifying the most divisive content and incentivizing sensationalism.[14]

It should be noted that algorithmic amplification may have differential effects on different populations. Survey data collected by the Pew Research Center show that most users report being exposed to a variety of viewpoints on social media.[15] Forty percent of social media users across different countries report being exposed to a diverse range of sources, according to data from a 2017 Reuters Institute Digital News Report.[16] A comprehensive review of the literature on political polarization and social media suggests that we need a more refined understanding of how echo chambers work and the mechanisms by which they can have an impact on users' process of radicalization.[17] In addition, findings on the impact of algorithms should be examined within the context of historical research on information consumption and individual and group behavior across other platforms, including television, radio and offline. While concerns around algorithmic amplification are relatively new, there is significant research across disciplines on related issues, including social, behavioral, communications, economic and other technology-related topics.

The underlying logic of algorithmic amplification may ultimately be traced to some platforms' core business model, which is to increase user engagement, extract data from users, and monetize that data and engagement through advertising or other transactions. Harassment, hate speech, and illegal content like child pornography and terrorist propaganda have higher engagement rates than more anodyne content.[18]

2. Platform responses to the harms posed by algorithmic amplification

2.1 Platforms have acted to mitigate harms posed by algorithmic amplification

Facebook's own researchers found in a 2016 internal report that "64% of all extremist group joins are due to our recommendation tools."[19] In one presentation in August 2020, internal Facebook researchers said roughly "70% of the top 100 most active US Civic Groups are considered non-recommendable for issues such as hate, misinfo, bullying and harassment".[20] In 2018, Facebook changed its algorithm to demote "borderline" content – harmful or distasteful content that did not

[14] Keach Hagey and Jeff Horwitz, "Facebook Tried to Make Its Platform a Healthier Place. It Got Angrier Instead," *The Wall Street Journal*, 15 September 2021: https://www.wsj.com/articles/facebook-algorithm-change-zuckerberg-11631654215
[15] Maeve Duggan and Aaron Smith, "The Political Environment on Social Media," *Pew Research Center*, 25 October 2016: https://www.pewresearch.org/internet/2016/10/25/the-political-environment-on-social-media/
[16] Nic Newman et al., "Reuters Institute Digital News Report 2," *Reuters Institute for the Study of Journalism*, 2017: https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital News Report 2017 web_0.pdf
[17] Joshua Aaron Tucker et al., "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature," March 19, 2018: https://ssrn.com/abstract=3144139.
[18] Anti-Defamation League, Avaaz, Decode Democracy, Mozilla and America's Open Technology Institute. "Trained for Deception: How Artificial Intelligence Fuels Online Disinformation". *Mozilla Foundation,* September 2021: https://foundation.mozilla.org/en/campaigns/trained-for-deception-how-artificial-intelligence-fuels-online-disinformation/
[19] Jeff Horwitz & Deepa Seetharaman, "Facebook Executives Shut Down Efforts to Make the Site Less Divisive," *Wall Street Journal*, May 26, 2020: https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499
[20] Horwitz, Jeff, "Facebook knew calls for violence plagued 'groups,' now plans overhaul," *Wall Street Journal,* January 31, 2021: https://www.wsj.com/articles/facebook-knew-calls-for-violence-plagued-groups-now-plans-overhaul-11612131374

quite violate its Terms of Service – as well as content deemed false by fact-checking organizations.[21] Before the 2020 US elections, the platform reported it had attempted to filter problematic groups, pages, and content from recommendations, reduce the distribution of borderline content, and add warning screens and fact checks to proactively prevent users from posting hateful content.[22]

Instagram also announced changes after January 6, 2021 to de-emphasize posts with bullying, hate speech, or the promotion of violence in users' Feed and Stories.[23] Instagram says it is exercising more care in what it recommends to teens and that it will nudge teens away from harmful topics.[24] Instagram has also given users more options about whether their feed reflects the platform's algorithm or reflects the chronological order of accounts they follow.[25]

Twitter rolled out policies in preparation for the 2020 U.S. elections, including not recommending tweets with warning labels.[26] After an internal review of their recommendation algorithms found greater amplification of right-leaning than left-leaning political content, Twitter stated that it would share aggregated data sets with outside researchers as part of its efforts to "reduce adverse impacts."[27]

In 2019, YouTube announced that it would reduce recommendations of 'borderline' content and content that could misinform users in harmful ways – in part by increasing human review and deploying machine-learning algorithms.[28] Comparative data showed this change resulted in fewer fringe channels shown alongside news videos during the 2020 US elections than during the 2016 elections, with spillover effects to Facebook and Twitter.[29] In response to a report that YouTube continued to amplify violent videos and misinformation, the platform reported it launched unspecified additional changes to reduce recommendations of harmful content.[30]

2.2 Platforms' Critique of External Algorithmic Harms Assessments

[21] Josh Constine, "Facebook will change algorithm to demote 'borderline content' that almost violates policies", *Tech Crunch,* 15 November 2018: https://techcrunch.com/2018/11/15/facebook-borderline-content/?guccounter=1
[22] Rosen, Guy. "Hate Speech Prevalence Has Dropped by Almost 50% on Facebook", *Meta,* 17 October 2021: https://about.fb.com/news/2021/10/hate-speech-prevalence-dropped-facebook/
[23] "How We Address Potentially Harmful Content on Feed and Stories", *Instagram*, 20 January 2022: https://about.instagram.com/blog/announcements/how-we-address-harmful-content-on-feed
[24] Adam Mosseri, "Raising the Standard for Protecting Teens and Supporting Parents Online", *Instagram*, 7 December 2021: https://about.instagram.com/blog/announcements/raising-the-standard-for-protecting-teens-and-supporting-parents-online
[25] Taylor Hatmaker, "Instagram's chronological feed is back", *TechCrunch, 5* January 2022: https://techcrunch.com/2022/01/05/instagram-chronological-feed/
[26] Vijaya Gadde & Kayvon Beykpour, "An update on our work around the 2020 US Elections", *Twitter,* 12 November 2020: https://blog.twitter.com/en_us/topics/company/2020/2020-election-update
[27] Ferenc Huszar et al., "Algorithmic amplification of politics on Twitter", *PNAS,* 119.1 (4 January 2022).
[28] The YouTube Team, "Continuing our work to improve recommendations on YouTube", *YouTube Blog,* 25 January 2019: https://blog.youtube/news-and-events/continuing-our-work-to-improve/
[29] Jack Nicas, "YouTube Cut Down Misinformation. Then It Boosted Fox News.", *The New York Times,* 3 November 2020: https://www.nytimes.com/2020/11/03/technology/youtube-misinformation-fox-news.html; Davey Alba, "YouTube's stronger election misinformation policies had a spillover effect on Twitter and Facebook, researchers say.", *The New York Times,* 14 October 2021: https://www.nytimes.com/2021/10/14/technology/distortions-youtube-policies.html
[30] Mozilla Foundation, *YouTube Regrets,* July 2021: https://assets.mofoprod.net/network/documents/Mozilla_YouTube_Regrets_Report.pdf; Clothilde Goujard, "YouTube's algorithm pushes hateful content and misinformation: Report", *Politico,* 7 July 2021: https://www.politico.eu/article/mozilla-firefox-report-youtube-algorithm-pushes-hateful-content-misinformation/

Researchers have used a variety of methods to conduct studies, including via simulations of online discourse. While independent accounts of the observed data have great merit, this mode of research has its limits. For example, these studies focus on a snapshot in time and cannot capture information flows over long periods.[31] Despite subsequent policy and product changes, platforms have contested the linkage between algorithmic amplification and online harms when that linkage was found by third parties. Researchers in turn do not have access to the corroborating data.[32] Access to data is key for conducting better research into algorithmic amplification, and studies have identified research gaps, a topic which will be covered in Workshop 3.

3. Potential voluntary measures to mitigate harms posed by algorithmic amplification

3.1 Reinforce knowledge and awareness of potential negative effects of algorithmic amplification

While media literacy initiatives have their challenges, and cannot be treated as a panacea, some believe that empowering users with more information via design changes or updated Terms of Services can be helpful in order to inoculate people from the potentially harmful effects of algorithmic amplification.[33] Proposed interventions include: telling users in clear terms how amplification works; explaining to users why specific content is shown to them; notifying users why a piece of content is demoted; and allowing users to better customize their feeds, for instance by introducing more granular controls that would allow them to adjust their likelihood of being exposed to certain types of "borderline" or "sensitive" content.

3.2 Introduce mechanisms to slow fast-spreading viral content.

Others have proposed that platforms could implement "circuit breakers" to stop the spread of harmful viral content, just as circuit breakers are used to stop the trade of securities when the market overheats. The trigger to stop spread could be based on the number of impressions or the rate of spread a given piece of content receives, with safeguards for the amplification of information the platform deems in the public interest.[34] Other frictive interventions include limiting the number of shares, requiring users to click through screens that seek to disrupt virality, asking users whether they want to share content that has been flagged, and implementing time delays for the transmission of certain content.

3.3 Legislative initiatives focusing on algorithmic amplifications

The following is a non-exhaustive list of legislative initiatives which seek to address some of the harms identified with algorithmic amplification. Proposals have been crafted to address algorithms through a variety of lenses, including privacy, transparency, competition, and law enforcement. Debate exists over the potential impact of such measures or whether similar initiatives have yielded meaningful results, analysis outside the immediate scope of this paper.

---

[31] Eli Lucherini et al., "Studying the societal impact of recommender systems using simulation", *Center for Information and Technology Policy,* 4 August 2021.
[32] Nicolas Kayser-Bril, "AlgorithmWatch forced to shut down Instagram monitoring project after threats from Facebook", *AlgorithmWatch,* 13 August, 2021: https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook/
[33] Monica Bulger and Patrick Davison, "The promises, challenges, and futures of media literacy," *Journal of Media Literacy Education* 10.1 (2018): 1-21.
[34] Ellen P. Goodman, "Digital Fidelity and Friction," *Nevada Law Journal*, 21: 2 (2021): 623-654.

- The proposed EU Digital Services Act (DSA) requires very large online platforms that use recommendation systems to outline in their Terms of Service the primary parameters used by algorithmic amplification systems. Companies must also explain how and to what extent users can control these parameters to adjust their platform experience. The DSA also provides that users should have access to an experience not based on profiling.[35]
- The DSA would require very large platforms to assess the systemic risks stemming from the functioning and use of their services at least once a year and take appropriate mitigating measures, including adapting the design and functioning of their algorithmic recommender systems so that they discourage and limit the dissemination of illegal content. They will also have to submit to external, independent audits.
- Finally, the DSA proposes a mechanism for facilitating data access to vetted researchers, including as regards algorithmic systems such as recommender systems.
- In the EU, the Platform-to-Business Regulation, as well as the New Deal for Consumers provide transparency on the general parameters for ranking systems to business users and consumers, respectively.
- The EU's Artificial Intelligence Act proposes a risk-based approach to AI regulation along a sliding scale of potential harms.[36]
- The General Data Protection Regulation sets rules on the profiling of individuals and conditions for automated decision making on the basis of such profiling.
- In the U.S., there are a number of proposed bills that would address algorithmic amplification, including the Filter Bubble Transparency Act[37] (requiring internet platforms to offer an alternative feed where content is not selected by "opaque algorithms" driven by personal data); the Protecting Americans from Dangerous Algorithms Act[38] (immunizing chronological but not other algorithmic ranking); the Algorithmic Justice and Online Platform Transparency Act[39] (prohibiting discriminatory use of personal information in algorithmic processes and requiring transparency in algorithmic decision making); and the Social Media NUDGE Act[40] (requiring study of algorithms and steps to reduce viral spread of harmful content) which further mandates the creation of "content agnostic" ways to create friction on platforms, to be codified and enforced by the Federal Trade Commission.
- The FTC has published a blog on the principles and best practices on the use of AI and Algorithms.[41]

---

[35] "The Digital Services Act package", European Commission, 1 February 2022: https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

[36] European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Acts, COM/2021/206, 2021

[37] Filter Bubble Transparency Act, S.2024, 117TH Cong. (2021): https://www.congress.gov/bill/117th-congress/senate-bill/2024/text?format=txt

[38] Protecting Americans from Dangerous Algorithms Act, H.R. 2154, 117TH Cong. (2021): https://www.congress.gov/bill/117th-congress/house-bill/2154/text

[39] Algorithmic Justice and Online Platform Transparency Act, S. 186, 117th Congress (2021): https://www.congress.gov/bill/117th-congress/senate-bill/1896/text

[40] Issie Lapowsky, "New bill would force social media giants to embrace friction – or else", *Protocol,* 10 February 2022: https://www.protocol.com/bulletins/social-media-nudge-act

[41] Smith, Andrew. "Using Artificial Intelligence and Algorithms," FTC Bureau of Consumer Protection, April 8, 2020. Using Artificial Intelligence and Algorithms | Federal Trade Commission (ftc.gov); Elisa Jillson. "Aiming for truth, fairness, and equity in your company's use of AI," April 19, 2021. Aiming for truth, fairness, and equity in your company's use of AI | Federal Trade Commission (ftc.gov)

4. Key Questions

1. What are the different types of risks and societal harms or concerns that are currently associated with algorithmic amplification?
2. What tools do various stakeholders (the public, journalists, researchers, regulators) need to mitigate/address the different risks relating to algorithmic amplification, recognizing that those risks may be different among different audiences (i.e., children) and different contexts?
3. What are the limitations of, and obstacles to, voluntary actions social media platforms and others have taken? And what improvements to these practices are contemplated, useful, and/or practical to mitigate risks?
4. What incentives could governments provide to improve both the analysis of this problem and potential solutions?
5. Additional perspective: Although this workshop is focusing on the *negative* effects of algorithmic amplification on social media platforms, algorithmic amplification may have effects for the platforms and their user bases that could be considered positive in context (e.g., connecting users based on common interests, providing users with non-harmful/problematic information they might have genuine interest in [and that continues to be non-harmful/problematic even if it is amplified to a broad user base], displaying content chronologically, etc.). To what extent should the positive effects of algorithmic amplification be considered when addressing its negative effects? What are the costs to the positive effects, if any, associated with addressing these negative effects, and how can negative effects be addressed while minimizing impact to the positive effects?