PANOPTYKON FOUNDATION

Not only content moderation: Creating rules for targeting content in the Digital Services Act or ancillary regulations

Introduction: Targeting as a source of power of online platforms

The business model of most online platforms (especially social media platforms) is based on the monetisation of personal data. Large numbers of users that platforms like Facebook or Google have managed to gather enable them to **accumulate vast amounts of users' personal data**, not only the data that users share on their own initiative, but – more importantly – metadata revealing users' actual behaviours, such as how they use the platform, what content they are exposed to and how they react to it.

This data is analysed **with the use of algorithms** and compared with other users' data for meaningful statistical correlations. The task of these algorithms is to predict users' characteristics that they have never consciously revealed, such as their psychometric profiles, IQ level, family situation, addictions, illnesses, beliefs etc., thus **creating detailed user profiles** which are then offered to advertisers and used for personalising content.

In order to continue gathering more and more data about users and making more and more money from targeted advertising, platforms need to **maintain user attention and increase the amount of time spent on the platform**. The dynamics that happen on the platforms can be summarised as a string of co-dependent events: the more time users spend on the platform, the more data they reveal; the more data they reveal, the more detailed their profiles are; the more detailed the profiles, the more revenues for platforms from selling targeted advertising.

To set this chain of events in motion and maximise the time spent on the platform, social media companies use algorithms to **amplify, rank, and target content** so that users are presented content that will be more engaging for them.

In summary, platforms control both the content that users won't see (moderation) and the content – paid or not – that users will see (targeting).

At the same time, since platforms operate as **walled gardens**, no one apart from the platforms has any insight into what personal data is used and what is the logic of the algorithms used for targeting. This insight is not offered even for advertisers who only commission a campaign and define characteristics of the audience they wish to reach, but do not have access to individual user profiles or control over which users and why the platform selects as targeted audience.

We view these issues as an underlying cause for many of the problems that the European Union wishes to address through the Digital Services Act or ancillary proposals.

Recommendations for the Digital Services Act or ancillary proposals:

1. The EU regulations should focus not only on content moderation but also on targeting, as it equally impacts fundamental rights.

While clarifying liability rules for content moderation is important for safeguarding freedom of expression, it is not the only usage of algorithms that has a crucial impact on fundamental rights. Targeting of content – be it sponsored or organic – is equally, if not more, important as it **determines what messages will be shown to users**, which ones will be amplified and which ones will receive less visibility. As opposed to content moderation which has equal effects for all users, algorithms used for targeting interpret users' personal data to determine which content will resonate better with them, thus creating a **risk of exploiting their vulnerabilities**, of which the most striking examples to date come from the use of data to target political adverts. This has grave implications for both the right to privacy and self-determination (as users do not have access to specific data that was used for targeting) and for the freedom of speech, as selective exposure to content can reinforce polarisation and close people in filter bubbles, thus negatively affecting the quality of public debate.

As surveillance-based business models rely on constant data collection and profiling, the platforms aim to maximise the time users spend on the platform by using ranking algorithms that promote content that is more engaging. It has a very negative impact on public health, especially for young people who become addicted to social media[1]. However, it is also detrimental to the quality of media and gives rise to clickbait and fake news, given that human psychology reacts more strongly to emotional or sensationalist messages.

2. From ad creation to delivery: Rules for sponsored content should cover the entire targeting process, not only the contents, funding, and reach of adverts.

In Europe, there is currently no clarity nor a common standard governing how much transparency is required into online platforms' advertising practices. This observation, calling for legal intervention and more harmonisation, is supported by the study commissioned by the European Commission in 2018[2]:

> *The main legal challenge, as apparent from the diversity of examples documented during the desk research, is that there is an abundance of disclosure practices fragmented across devices, jurisdictions and providers, while the legislative framework is open as to how and how much disclosure must be provided.*

We welcome the proposal of the JURI committee in the draft DSA report[3] to introduce mandatory ad archives for all adverts and the IMCO committee's[4] due diligence requirements for

---

[1] See for example: Y. Hou et al., *Social media addiction: Its impact, mediation, and intervention*, https://cyberpsychology.eu/article/view/11562/10373

[2] https://ec.europa.eu/info/sites/info/files/osm-final-report_en.pdf

[3] https://www.europarl.europa.eu/doceo/document/JURI-PR-650529_EN.pdf

[4] https://www.europarl.europa.eu/doceo/document/IMCO-PR-648474_EN.pdf

advertising. However, transparency and due diligence requirements related to advertising should not be limited to the verification of advertisers and contents of the adverts, and clear labelling[5] or present only the effects of targeting[6] but **should cover the entire targeting process: from ad creation to delivery**.

On its own targeting poses significant risks for user privacy and can lead to discrimination as it exploits (potentially sensitive) information about people in determining which users should see particular adverts. Moreover, in the era of personalised advertising **targeting is not only essential to delivering adverts but it might even determine the content of the advert**, as the message itself might be tailored to the target audience the advertiser wishes to reach. Facebook already offers the service of adjusting the content of the ad to the characteristics of users that the advertiser wishes to reach[7].

The need for regulating this issue in the DSA is further justified by evidence from multiple studies[8] that voluntary transparency measures implemented by online platforms did indeed significantly increase the transparency of the contents of adverts but are **superficial and not sufficient** to enable verification of whether users' vulnerabilities are exploited in the targeting process, either by advertisers or by the platform. Therefore, concrete requirements related to targeting should be put forth in binding regulation. **We elaborate on that in point 4.**

3. Beyond political advertising: A need for transparency of all adverts.

As proposed by the JURI committee in the draft DSA report, **transparency requirements for targeting should apply to all ads**, not only to political ads. This is due to two reasons: first, the concept of a "political" ad is highly ambiguous and difficult to enforce in practice (of which there is plenty of evidence[9]), and second: commercial advertising online also creates many risks that may require intervention - especially on behalf of children and other vulnerable groups.

For researchers and institutions scrutinising political adverts, transparency of targeting for all adverts avoids the problems in agreeing on a controversial definition of "political issues." Secondly, it is the only way to make sure that researchers and institutions are not missing anything, and that platforms' rules are being enforced properly.

Comprehensive transparency rules also enable research into harmful commercial advertising, and are important from the perspective of consumer protection and data protection authorities. Examples of harms include the sale of harmful, illegal, or regulated products, the use of

[5] As proposed by the IMCO committee.
[6] As proposed by the JURI committee.
[7] Facebook: Dynamic Ads, https://www.facebook.com/business/help/397103717129942?id=1913105122334058
[8] See for example: ERGA report on the implementation of the EU Code of practice on disinformation: http://erga-online.eu/?p=732, empirical academic research into Facebook's explanations: https://www.researchgate.net/publication/323248639_Investigating_Ad_Transparency_Mechanisms_in_Social_Media_A_Case_Study_of_Facebook's_Explanations, Panoptykon Foundation: https://panoptykon.org/political-ads-report, European Partnership for Democracy: http://epd.eu/virtual-insanity.
[9] https://dl.acm.org/doi/fullHtml/10.1145/3366423.3380109, https://github.com/citp/mistaken-ad-enforcement/blob/master/estimating-publication-rates-of-non-election-ads.pdf, https://news.sky.com/story/researchers-fear-catastrophe-as-political-ads-disappear-from-facebook-library-11882988

manipulative or deceptive sales tactics, and discriminatory targeting of ads. A comprehensive ad library would help to research these practices and hold them accountable.

4. Scope of transparency: Insight into the parameters selected by advertisers and the optimisation process controlled by the platform.

Even though the targeting process is initiated by advertisers who create ad campaigns and choose the profile of their target audience, the **key role still belongs to online platforms**. While advertisers do make their own choices, their choices have been shepherded by platforms and increasingly rely on data that was collected or inferred by them.

This means that platforms are not merely passive intermediaries between advertisers and users. Their algorithms interpret criteria selected by advertisers and deliver ads in a way that fulfills advertisers' objectives. Online platforms, functioning as walled gardens, **offer advertisers the means to reach very specific groups without having to collect data**. Moreover, empirical academic research into Facebook's advertising machinery confirms that optimisation algorithms can lead to discrimination of certain groups even when the advertiser did not intend to do that[10], as well as contribute to political polarisation and creating filter bubbles by showing users only these political adverts that correspond with their views[11].

It is also important to distinguish between information that should be publicly available in ad archives and information that should be revealed individually to the user.

Therefore we propose the following transparency framework to be mandated in the DSA:

- Public disclosures made by online platforms in relation to advertising should:
    - **cover decisions made by both actors in the targeting process**, i.e. the advertiser (both in terms of selection of the audience and determination of the campaign strategy, including what was it optimised for), as well as methods and parameters used by the platform in the optimisation process;
    - for decisions made by the advertisers - include an **equivalent level of information as is offered to the actual ad buyer** when commissioning the campaign (e.g. actual audience demographics, location and other targeting criteria);
    - with regard to the platform's role - **explain the prediction model (including fairness criteria and variables)** used to achieve the optimisation goal (without revealing the source code).
- Individuals should have access to:
    - a **personal ad library** that would cover both all ads that have been targeted at them and all advertisers who targeted them within specified time (we propose 5 years). Because it would show only ads that are relevant to an individual user, this interface would not be as overwhelming as the public ad library, which

[10] M. Ali et al., *Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes*, https://arxiv.org/pdf/1904.02095.pdf
[11] M. Ali et al., *Ad delivery algorithms: The hidden arbiters of political messaging*, https://arxiv.org/abs/1912.04255

presents a vast amount of ads and requires advanced research skills, thereby offering more meaningful transparency towards the user;

- **information explaining the logic behind targeting of specific ads**. At a minimum, this individual explanation should enable the user to understand why they are seeing a given ad and should include data that is specific, accurate and relevant to the user, such as:
  - All parameters selected by the advertiser which are relevant to the user;
  - Reasons why the platform has decided that the user meets the advertiser's criteria, including personal data that were relevant in this process;
  - Specific source of user's data (e.g. website tracking, mobile app, newsletter, loyalty card) if data was uploaded by the advertiser;
  - Optimisation goal selected by the advertiser and concrete reasons why the platform has decided that the user should be reached with this ad.

5. Beyond transparency: Effective control tools for users and regulators.

Currently users do not have the full control over their data that is collected by online platforms or inferred about them. Neither can they verify true reasons for being included in a particular group of users that the advertiser wanted to reach. In fact, leading platforms enable users to verify only a short list of uncontroversial interests that platforms are willing to reveal. This is by no means an exhaustive list of the results of constant behavioural observation and algorithmic analysis that we've described in the introduction. As documented by extensive research, including facts revealed by whistleblowers in the Cambridge Analytica scandal, **characteristics assigned to users can be sensitive or reveal users' vulnerabilities**.

At the same time, we acknowledge the difficulty in explaining and giving users control over the ad optimisation process which uses machine learning and complex algorithms. Even in the absence of sensitive data, individual characteristics can still be revealed in a dataset because of so-called **"proxy variables"** — elements of the dataset that are closely related to sensitive attributes. For example, the number of hours spent daily on a platform may be a proxy for age or employment status, and the set of liked pages may be a proxy for gender, sexual orientation, or political opinion.

This phenomenon becomes even more pronounced when machine learning algorithms are used to select target groups, since algorithms excel at combining multiple proxies to find patterns equivalent to sensitive attributes. An algorithm may therefore recognize individuals as similar, and treat them as such, without labelling them as "male" and "female" or "liberal" and "conservative." This makes it challenging to explain with confidence what factors had the most decisive impact on generating an ad recommendation for a particular individual.

However, these complexities should not justify the opacity of predictive models and hinder the possibility for users to effectively control their data.

We have the following recommendations:

- with regard to access to data and user settings:
    - Users should have **full access to their data**, including inferred representations, regardless of the purpose for which they were generated. Users should have an easy option to delete or correct every piece of inferred or observed data.
    - **By default** users' behavioural data should not be used for targeted advertising and, in particular, for political advertising. Users should have the possibility to **opt-in** for targeting (separately for political advertising and separately for other adverts).
    - To mitigate consent fatigue and the burden of managing their data, users should not be bombarded with requests for consent and default settings should be set to 'no targeting'. Any change in these settings should **require action on the user's own initiative**.
    - Users who would like to actively manage their preferences should be able to communicate their choices in a standardised way, e.g. via their browser or via an open, programmable, **personal API** that should enable connection with the user's own client (such as a trusted data management service).
- with regard to accountability of algorithms used for ad targeting:
    - platforms should produce thorough **documentation of their models**, including fairness criteria for their ad optimisation, and allow for external audits;
    - regulators should be able to **verify and challenge algorithmic design choices** that involve information about individuals.