EUROPEAN PARTNERSHIP FOR DEMOCRACY

Targeting and amplification in online political advertising

The online ad ecosystem is driven by two processes: targeting – advertisers picking who should see their ad – and amplification – platforms' algorithms picking a select audience from within this targeted group. While these are two parts of the same process, they are very different in nature and pose different challenges when used in political advertising.

Yet the European Commission's proposal for a Regulation on the transparency and targeting of political advertising (the 'Regulation' hereinafter) does not clearly define both processes, nor does the Commission differentiate between both processes in the proposal for a regulation. Adding to this is the systemic failure by authorities to enforce the GDPR, which poses policy dilemmas that must be addressed in this Regulation, but are not tackled in the proposal.

In this paper, we present the proposed Regulation of the use of data in political advertising and advocate for an alternative approach that considers 'targeting' and 'amplification' separately and privileges a pragmatic understanding of the GDPR over a legalistic one. This paper is the second of a series of policy briefs on online political advertising and the regulatory proposal.

Background: The use of data in online advertising

Virtually all online advertising engages in **behavioural targeting**, meaning that it is directed at defined audiences tailored by the sponsors and publishers of the ads. The large majority of online ads are targeted to individuals based on the personal data collected about them by the AdTech industry. Before discussing how **personal data** is used in political advertising and how to regulate its use, it is worth differentiating between the different types of data, both depending on **what they reveal about the individual and the method used to obtain it.**

First of all, some data is **'special category data'** as per the GDPR if the information that the data reveals about a person reveals "racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation." This subcategory of personal data can only be processed under certain conditions – including explicit consent for a specified purpose or purposes.

Secondly, depending on the **methods** used to obtain data, we distinguish **provided**, **observed**, and **inferred** data, following the Guidelines by the European Data Protection Board.

Provided data (also known as revealed or declared data) is information actively provided to the data collector by the individual with their consent, as defined in Article 4 (11) GDPR. This could be the location, age, or gender you provide when setting up an Instagram account, for instance. Provided data is only a minor fraction of the data used in online advertising, and lots of the data individuals consent to reveal is not revealed deliberately, but in flawed ways that involve the use of dark patterns. The use of dark patterns also affects the collection of other types of data.

Observed data (also referred to as tracking-based data) includes all data that can be observed from from a person's activities on an online service. In other words, this is the data passively provided by an individual. This includes, for instance, the videos watched by a person on

YouTube, the reactions to content on social media or the geo-location.

Inferred data (also referred to as derivate data) is qualitatively different from both provided and observed data, as it is not actively or passively obtained from the individual, but created based on provided and observed data. Inferred data is generated by algorithms.

It is worth dividing inferred data between 'human understandable' data and 'non-human understandable data'. **Human understandable inferred data** results in the creation of profiles that can be understood and analysed by humans. Examples include Meta profiling users who like posts by Greenpeace as interested in 'Environmentalism' and Google identifying that an individual is most likely a woman between 18 and 24 because of the videos she watches on YouTube.

In these cases, data inference is used to label individuals and build personal advertising profiles. It must be noticed that by generating this type of inferred data, the Adtech industry profiles individuals into groups that largely overlap with racial, ethnic, political, religious and/or sexual orientation groups, making such data a de facto proxy for **special category data under the GDPR** (see below). Even this less sophisticated form of **inferred data use is highly intrusive** because the 'interest' labels AdTech industry attributes to individuals go as far as 'Incest/ Abuse Support', 'Paranormal Phenomena' or 'Epilepsy'.

More sophisticated forms of inferred data are not understandable to humans. This subtype of inferred data is generated combining greater amounts and types of data and using more powerful machine-learning algorithms – in processes opaque to the data controllers themselves. Examples include the case uncovered by a case study on Facebook's algorithms, which showed that the use of inferred data exploits individual's vulnerabilities, including those associated to health issues with neither control nor specific knowledge by the data controller. Due to the complexity and great difficulty to validate and trace inferences in such cases, it is hardly possible to determine if the data generated by the use of machine-learning algorithms should classify as special category data or not. This raises questions on whether the GDPR has an answer to AI-related data protection issues, as acknowledged by the EPRS.

It is worth highlighting that not all online advertising engages in these intrusive practices. There is a small minority of advertisers using an innovative privacy-respecting technique called **contextual advertising.** Contextual advertising relies on the content of the website that an individual is visiting to target ads, instead of the traits (be them provided, observed or inferred) of the individual.

The use of data in the proposed Regulation

The use of data in online political advertising is regulated in Chapter III of the Regulation, particularly Article 12 on the specific requirements related to targeting and amplification. The proposal defines targeting and amplification synonymously as: "techniques that are used either to address a tailored political advertisement only to a specific person or group of persons or to increase the circulation, reach or visibility of a political advertisement."

First, the proposal restricts the possibility to use **special**

Provided data	proactively revealed data
Observed data	passively revealed data
Inferred data (human understanable)	data generated by algorithms based on observed and provided data; human explainable
Inferred data (non-human understandable)	data generated by algorithms based on observed and provided data; not human explainable

category data in political advertising, unless in two specific cases: (a) either the individual has given explicit consent for this specified processing purpose, or (b) the data is processed by a controller who is an organisation of which the individual is a member of (i.e., political party membership). Yet a CSO has identified that, in the current context where large platforms can easily obtain consent in flawed and dubious manners, the exemptions in the proposal (Article 12.2) are not exemptions, but the rule.

Second, the proposal establishes that the controllers of data used in online political advertising must fulfil **three requirements:**

1. a adopt and implement an internal policy describing clearly and in plain language the use of targeting and amplification techniques, and retain such policy for a period of five years;
2. keep records on the use of targeting or amplification, the relevant mechanisms, techniques and parameters used, and the source(s) of personal data used; and
3. provide through transparency notices additional information so that the individual concerned can understand the logic and the main parameters of the technique used, and the use of third-party data and additional analytical techniques.

Information on the use of personal data must be included in ad transparency notices by publishers that make use of targeting or amplification techniques, alongside a reference to effective means for targeted individuals to exercise their rights under GDPR. When the data controller is different from the advertising publisher, the controller shall transmit the internal policy or a reference to it to the political advertising publisher.

Redefining targeting and amplification

A shortcoming of the Regulation is that it **does not define 'targeting' and 'amplification' techniques separately**. By considering them as a single type of technique, the impact of their use is not properly accounted for, and it is not possible to devise adequate, nuanced policy responses. Below, we **explain the difference between targeting and amplification techniques**.

The use of data to tailor the audience of a political advertisement happens in two phases. First, targeting takes place when sponsors define the potential audience of an ad. Second, amplification takes place when the publisher's machine learning algorithms determine the individuals within the potential audiences who actually receive the ad - the actual audience of the ad. Amplification is also known ad optimisation or ad delivery.

Targeting

Targeting is the technique available to the sponsor of the ads to determine who they would like to reach with an ad. The potential audience can be defined based either on the content of the website they are visiting or the personal traits, such as demographics or interests.

A political party can choose to run an ad over a platform (for instance, Facebook) that will only be seen by those users of Facebook who have been identified as part of a subgroup. The political party might choose to target the ad at men between 25- and 30-years old living in the Region of Brussels, for instance.

When large amounts of data are collected and further inferred data are generated by platforms' algorithms and combined in a way which reveals sensitive traits of the individuals, this enables political messages to be tailored to hyper-specific audiences. This poses an obvious risk of manipulation of electoral processes and the public space in general. The most well-known case of abuse of data in targeted online political advertising is the Cambridge Analytica Scandal, where a PR consultancy collected and processed massive amounts of personal data to profile Facebook users along psycho-graphic lines, and then used Facebook's 'custom audience' tools to target them in the context of the Brexit Referendum, amongst others.

Amplification

Amplification starts where targeting ends. Once the sponsor of an ad has defined its target audience, amplification techniques determine who within the potential audience will actually see the ad. The publisher of the ads (Google or Meta platforms) optimises the use of the sponsor's ad budget to select the most relevant recipients of the ads based on the processing of immense amounts of data by powerful artificial intelligence to

auction in real time ad views and clicks among different sponsors.

As **the aim of the platforms is to maximise their profits derived from advertising**, the use of amplification techniques for political advertising **has inherent and undesirable by-products including the creation of filter bubbles, the fostering of polarisation and the fragmentation of the public space of deliberation.** Based on the content of the ad and massive amounts of inferred date, the ad delivery algorithm selects within the targeted audience those individuals most likely to react to the ad in specific ways, such as clicking, liking, sharing or watching a given ad. In this way, platforms need to deliver ads to a smaller audience to provide the same value to the sponsor of the ad, such as clicks, full views of videos and/or likes.

Following the example above, the ad delivery algorithm of Facebook would select those men between 25- and 30-years old living in the Region of Brussels who are most likely to be react to the ad in specific ways, such as clicking on the ad or sharing it based on data about them generated by machine learning algorithms which is not human-understandable.

The lack of researcher access to information on the algorithms and the inherent deficit of explainability of the algorithms makes it difficult to fully assess the impact of ad delivery algorithms and the data they generate and use. However, even without access to the internal logic of the algorithms, it is possible to analyse their outputs and outcomes. The observed outcomes are that algorithms discriminate among audiences along the lines of special category data (such as gender and ethnic identity), which in the case of political advertising creates risks for the integrity of public debate and electoral processes, including **the creation of filter bubbles, polarisation and price discrimination for sponsors of political ads.**

A group of scholars ran an experiment that showcased how amplification techniques operate on Facebook. They measured how Facebook delivers ads to different groups, depending on an ad's content (e.g., the political viewpoint featured) and targeting criteria. They found that platforms' ad **delivery algorithms** (amplification techniques) selectively deliver ads within these target audiences in ways that lead to **demographic skews along race, political alignment and gender lines**, often without the knowledge of the sponsor.

This effect is most acute when advertisers use small budgets, as Facebook's delivery algorithm tends to preferentially deliver to the users who are, according to Facebook's estimation, most relevant. Studies have shown that budgets of political ads are normally very low in Europe – often under EUR 100. This is problematic as **political parties and candidates across Europe are paying different prices** to reach audiences over online platforms in Europe depending on the political preferences of the users they target and reach.

The research also found that Facebook's ad delivery algorithms effectively differentiate the price of reaching a user based on their inferred political alignment with content of the ad, **inhibiting political campaigns from reaching voters with diverse political views.**

As **ad amplification** involves the processing of large amounts of data, it **cannot be apprehended by the individuals who sees the ads**, even if the Regulation indicates that such transparency is required 'with the same level of detail as used for the targeting' and 'in user-friendly (…) plain language.' Yet for ad amplification, those two legal requirements are at best mutually exclusive and, therefore, **the transparency obligations for amplification can be deemed as impossible to comply with**.

In other words, targeting can be restricted and made transparent, whereas amplification techniques are inherently opaque and rely on the use of massive amounts of real-time data. Targeting poses threats to the integrity of electoral processes when the type and amount of data processed in targeting is not restricted, whereas an inherent by-product of amplification techniques is the creation of filter bubbles, polarisation and price discrimination for sponsors of political ads.

GDPR interpretation & implementation and political advertising

Before considering how to best regulate the processing of data for political advertising purposes, it is necessary

to discuss the main data protection rules which the Regulatory proposal builds upon. In particular, it is necessary to discuss the interpretation of the GDPR in relation to inferred data and the enforcement of the GDPR.

GDPR interpretation & inferred data

There has been a debate on the interpretation of the GDPR when machine learning AI is used to process data. There are three possible interpretations of the relation between the GDPR and the technology used in inferred data, which are not fully mutually exclusive.

First, some conclude that AI can be GDPR compliant, if the obligations to provide information to the data subject are read narrowly and if platforms ensure that special category data can be inferred from non-special category data. Second, others indicate that AI-generated data cannot be protected unless a new right to 'reasonable inference' is included to the GDPR. Third, it has been argued that, due to the very nature of machine-learning AI-generated data, the data subject cannot possibly receive meaningful explanations on the logic of processing regardless of the transparency obligations imposed on the data controller, as AI data generation algorithms are designed to be black boxes. Those arguing in this line warn of a 'transparency fallacy' and that data protection law that mandates 'transparency by design' would be needed to protect the rights of data subjects established in the GDPR.

The debate on the interpretation of the GDPR is likely to evolve as the European Data Protection Board (EDPB) and the European Court of Justice (ECJ) eventually come to a common and consolidated understanding. In the meantime, there will be legal uncertainty on the interpretation of the GDPR on AI-generated data, giving raise to the risk that the spirit of the GDPR is not respected. This leaves **individuals' vulnerabilities exposed, allowing malign political actors and irresponsible economic operators to exploit these vulnerabilities.**

Limited GDPR enforcement

While the regulatory proposal on political advertising builds in part on GDPR enforcement, the enforcement regime of the GDPR has shown serious shortcomings since its entry into force. There is ample evidence of systematic GDPR under-enforcement that affects the use of data in online advertising. For instance, Politico Europe recently reported that the Irish DPA 'lobbied to allow social networks to bypass user consent requirements within EU privacy rules'. In addition, three years after the entry into force of the GDPR, Slovenia is yet to adapt its national data protection framework. Moreover, the European Ombudsman has opened an inquiry into how the European Commission has been monitoring the application of data protection regulations in Ireland, calling into question the willingness and capacity of the Commission to act as the 'Guardian of the Treaties' regarding the protection of personal data as enshrined in Article 16 TFEU.

In a 2021 resolution on the Commission evaluation report on the implementation of the GDPR, the European Parliament acknowledged the "uneven and sometimes non-existent enforcement of the GDPR by national DPAs" and stressed the need for better enforcement on online advertising, micro-targeting, and algorithmic profiling. It even expressed the concern that the enforcement has not substantially improved compared to the situation under the Data Protection Directive of 1995.

CSOs have reached even more worrisome conclusions, as they have documented that DPAs have given up on enforcing a series of GDPR obligations for online advertising service providers. In this context, the AdTech business model has evolved to make use of data breaches that occur on a daily basis, counting on the DPAs' lack of capacity or willingness to enforce any compliance with the principles of data minimisation and purpose limitation. As some CSOs have highlighted, the very use and abuse of inferred data can be regarded as a circumvention to the protection of the data subject under the GDPR, as individuals who share their data are deprived from full consent, control, portability and protection for the processing of personal data.

Conclusions on GDPR & political ads

This situation poses a policy dilemma in terms of

regulating the specific case of the processing of data in political advertising when the general Regulation is not effectively implemented. One must either expect the enforcement of the GDPR to have improved substantially by the time that the Regulation enters into force, or account for the GDPR shortcomings in the political advertising Regulation.

In the case of improved enforcement of GDPR, it might be reasonable that political advertising regulation would simply have to enhance transparency as proposed. In the scenario of continued insufficient enforcement of GDPR, it would be advisable to impose restrictions on the processing of data for political advertising.

Even if we expect that the enforcement of the GDPR will eventually improve following the adoption of the DSA, in the best case scenario this will improve slowly over time. For the GDPR to guarantee the protection of the data subjects of inferred data, the academic and policy debate suggests the GDPR would need to be reviewed, upgraded and complemented by other legislative initiatives – even in the most ambitious reading of the GDPR.

In the meantime, legal uncertainty and inadequate enforcement of the rules on the use of data in political advertising will likely mean that yet another cycle of European elections takes place without adequate safeguards on the use of data in political campaigning. Under these circumstances, we can expect malign and savvy political actors to abuse personal data in advertising in **a scandal similar to Cambridge Analytica Scandal.**

Regulating targeting & amplification coherently

In this paper, we have argued why the approach to the use of data in political advertising in the Regulation on online political advertising would benefit from amendments that take stock of (a) the differences between targeting and amplification techniques, (b) the state of enforcement of the GDPR and (c) the ambivalence of the interpretation of the GDPR concerning AI-generated data.

Below, we suggest broad ideas for amendments on the use of personal data in political advertising, that advance the integrity of and trust in electoral processes and reinforces the European framework of human rights.

Together, these policy recommendations can advance the integrity of elections and fundamental freedoms, regardless of the success of the enforcement and interpretation of the GDPR in the near future, or technological developments. Adopting the proposed recommendations would result in simple, easy to monitor rules, that would prevent abuses on the

EPD proposal for limiting data use for targeting and amplification of political ads

	Targeting	Amplification
Provided data	✓	X
Observed data	X	X
Inferred data (human understandable)	X	X
Inferred data (non-human understandable)	X	X

processing of personal data, lead to a less fragmented and polarised public space of deliberation, and advance trust in democratic processes by the European citizenry.

That being said, the proportionality of the recommendations still hinges on the amendments to the definition of online political advertising and to the transparency over ad repositories to be extended to political ad publishers that are not covered by the DSA. The recommendations are particularly necessary for capturing core political actors, such as parties, candidates and anyone receiving compensation for campaigning on their behalf. However, they may be excessive if the definition remains excessively broad, capturing the legitimate awareness-raising and fundraising activities of civil society and allowing for arbitrary operationalisation and implementation by online platforms.

The extent of negative impacts on democratic processes of the use of data in targeting – such as the fragmentation of public debate and harms to right to receive information – depend in turn on whether watchdogs and the public can easily access complete, real-time information on the use on data in political advertising, as any risk associated to the use of data would be substantially mitigated by a mandate for transparency for all political ads over ad repositories.

We recommend co-legislators to consider the following policy recommendations regarding Chapter III of the Regulation:

- **Include a ban the processing of all observed and inferred data in political advertising, for both targeting and amplification.**

  Machine learning-generated inferred data poses the greatest threats to privacy and democratic processes due to its complexity and opacity, and is mired in legal uncertainty regarding its use. Less sophisticated forms of inferred data and to some extent observed data also pose threats to privacy and electoral processes as the line between special category data and non-special category data is not clear for these types of data. In the context of poor GDPR enforcement the risks associated to the processing of these types of data are multiplied. Similar considerations underpin the opinion of the European Data Protection Supervisor on the proposal for a Regulation, which calls for the strongest possible restrictions on the use of data in political advertising, this is, a full ban on political microtargeting.

- **Restrict options available for the targeting of political ads to revealed data, including age, language, general location and possibly some other provided identity features or declared interest categories. Impose a ban the use of revealed data in amplification, effectively banning amplification techniques for political advertising.**

  Such targeting allows for political parties to campaign for local elections in the language that is most relevant to their audience, while excluding the many pernicious effects of targeting we see today. This is compatible with both the right to information of individuals not targeted over ad libraries and with freedom on the means of expression. This restriction is the status quo (following self-regulation) for political ads published on Alphabet platforms and could be mandated as the maximum level of granularity for all platforms, creating a level playing field on the use of data.

  Any restriction on the use of data in targeting and amplification – such as those proposed in this paper – might be understood as a restriction on freedom of the means of expression. However, the negative impact on freedom of expression as freedom on the means of expression must be regarded as minimal, proportionate and necessary for a public legitimate interest, considering that freedom of expression is a fundamental liberty but not an absolute one. This means that it must be weighed against the fundamental rights to information and protection of personal data, which are also enshrined in the Treaties and the Charter of Fundamental Rights of the European Union and the right to free elections established in the European Convention on Human Rights.