Benchmarks
for Evaluating Online Political Ads Definitions

Vera Sosnovik
Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG
France

Oana Goga
LIX, CNRS, Inria,
Ecole Polytechnique, Institut Polytechnique de Paris
France

1 INTRODUCTION

The political campaigning landscape has changed significantly with the digitalisation of our public sphere, which has created new opportunities for political participation, but also significant risks to the integrity of elections and the political debate. The lack of transparency of which ads are shown to whom, why, and who paid for them, further creates a situation where anyone - from a political party to a foreign advertising firm like Cambridge Analytica - can distort the political debate and easily evade public interest scrutiny.

To be able to monitor sponsored political ads social networking services try to regulate political advertising. For example, Facebook allows political ads, but advertisers have to verify their accounts and have to self-label their political ads and advertisers can only send political ads in the country they reside. Google allows political ads, but advertisers can use only geographic location, age, gender, and contextual targeting to target political ads. On the contrary, Twitter and TikTok ban all political advertising altogether.

Many of these measures are positive developments, but *they assume we can reliably detect political ads.*

This is not the case as even what constitutes political speech is under heavy debate. At the moment, each ad platform has a different definition for political ads, and the European Commission, which is trying to regulate online platforms through the Digital Services Act – DSA [2] and the European Democracy Action Plan – EDAP [1], is currently gathering working definitions of political ads from civil societies for a deeper analysis. While many agree that ads from political actors and ads about elections and voting should be considered political, the debate turns around on how to handle social issue ads such as climate change or refugees that might only indirectly influence legislation and elections and might not come from traditional political actors.

Currently, every platform and government has a different definition of what constitutes digital political advertising. and, "*what is a good definition for political advertising?*" is still an open question. In this paper, we contribute to this debate by proposing a set of practical benchmarks for evaluating definitions of political ads that allow us to compare them across various dimensions. Provided a definition and a set of ads, our benchmarks test. We assess the quality of the four definitions of political advertisement from different sources (from social media platforms and official government documents).

2 BENCHMARKS

We proposed the following benchmarks for the political ad definitions.

(1) *agreement* – do users agree on what ads are political and which ones are not?
(2) *influence* – is the definition able to capture ads that can influence people's voting behavior?
(3) *divisiveness* – is the definition able to catch ads that are divisive across different racial, age, and gender groups of people
(4) *humanitarian aid* – is the definition able to distinguish between advocacy ads on different social issues that try to influence opinions and legislation and operational ads that only try to mobilize users to help people in need.

2.1 Data collection and experiments

We assess the quality of the four definitions political advertisement: Twitter definition [10], Facebook definition [5], European Parliament definition[6] and European Commission[3].

To assess the quality of the definitions, we set up a series of experiments. We downloaded all Meta Ad library ads published during July 2022 and randomly selected 500 unique advertisements.

To analyze it, we set up surveys on Qualtrics where for each ad, we ask respondents questions about the ad's message [8]. We hired workers through Prolific, and we redirected them to fill out the survey [7]. Each worker answered questions about 25 random ads from the pool of 500 ads.

In the first study, we asked workers to label ads according to the definition. We did a separate survey for each definition.

In the second study, we asked workers three questions:

- Do you think, through this message, **the advertiser intended** to influence a legislative or regulatory process or voting behavior at the national, regional, local, or at political party level, and their outcome?

- Do you think **this message** could influence (with or without a direct intensity of an advertiser) a legislative or regulatory process or voting behavior at the national, regional, local or at political party level, and their outcome?
- Do you think this ad is divisive across different ethnic, social, and age groups of people?

We repeat these studies for 100 random ads with unique text from IRA [4]. *3,517 Facebook ads that were reported to have been bought by the Internet Research Agency were released by the House Intelligence Committee as a set of redacted PDF files and parsed by irads [9].*

In the third study, we created a survey with the following two questions:

- Does the message of the ad have an **operational focus**, such as encouragement to participate, donate or promote a development project or humanitarian aid?
- Does the message of the ad have an **advocacy focus** such as promoting certain causes by persuading citizens and state actors into promoting and adopting certain public policies across different areas such as the economy, election systems, environmental politics, or law?

Since the third study was only about social issues ads, we did a pre-selection of advertisements. We selected 100 unique ads published by only non-governmental and charity organizations.

3 RESULTS

We then launched our studies on Prolific [7], where we redirected workers to the Qualtrics surveys. Each worker answered questions about 25 random ads from one of the datasets.

The first study's results are shown in the table 2. None of the definitions manages to achieve agreement among workers of more than 60%. The Twitter definition has the highest percentage of ads that workers agreed on. Conversely, Facebook has the lowest amount of advertisements that do not cause disagreement. Twitter does not include issue ads into the definition and only focuses on ads that directly connect with political actors or elections. On the opposite, official Facebook definition includes ads about social issues. This could be on of the possible reason for the lower percentage of agreement. On the opposite, the official Facebook definition includes ads about social issues. This could be a possible reason for the Facebook definition's low percentage of agreement. The European Commission's definition slightly outperforms the amendment that European Parliament proposed.

We evaluate definitions by second and third benchmarks in the second study on two datasets (tables 3, 4). All four definitions performed well on the Facebook dataset. All of them were able to catch more than 80% ads that could influence people's voting behavior, and around 80% and more advertisements that are divisive were detected as well. However, the results are different from the IRA dataset. While the results of European Commission and Facebook definitions didn't drop, Twitter and European Parliament's definitions' performances significantly decreased on the IRA dataset. These definitions only detected more than 50% of problematic advertisements.

The fourth benchmark's results are present in the table 5. European Parliament and Twitter definitions outperformed others in the ability not to label operational ads as political. However, they were able to catch only 59% of advocacy ads. On the opposite, with the European Commission definition, more than 70% of advocacy ads were labeled as political, but it shows the worst performance with operational: 19% of them labeled as political. Facebook definition shows the most stable results on the fourth benchmark with mislabeling 9.5% of operational digital ads and catching 62% of advocacy ads.

Facebook definition outperformed others in the second and third benchmarks. It also showed the most stable result in the fourth benchmark. However, this definition has the lowest agreement among annotators. On the opposite, Twitter definition, while having the highest agreement, has the lowest result in other benchmarks. It shows that considering ads about social issues as political helps to catch problematic ads that are divisive and can influence people's voting behavior. Nonetheless, these ads create a more significant disagreement among workers and require a more detailed description.

4 CONCLUSION

The growth of political advertising and its misuse has led to social media platforms and the government imposing restrictions on them. However, they are still determining what political ads are. To be able to choose a proper definition, in this work, we propose four benchmarks for the evaluation quality of a political advertising definition. We assess the quality of two definitions proposed by social media platforms and two definitions from governmental organizations. We find that considering social issue ads as political increases the ability of a definition to catch divisive ads and ads that can influence people's voting behavior. However, this type of advertisement seems to be the most confusing for workers who labeled sponsored political content.

REFERENCES

[1] European commission. 2020. *Public consultation: Commission seeks citizens' views in preparation of new European Democracy Action Plan.* https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1352

[2] European commission. 2021. *The Digital Services Act package.* https://ec.europa.eu/digital-single-market/en/digital-services-act-package

[3] European commission. 2021. *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the transparency and targeting of political advertising.* https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1352

Dataset's name	All ads	Adver.influence	Mess.influence	Divisive	Operational†	Advocacy
FB dataset	500	362	379	205	-	-
IRA ads	100	56	63	59	-	-
NGO ads	100	-	-	-	21	71

Table 1: Datasets' description. Operational† means that an ad is only operational and does not have advocacy label. Advertisements that are both operational and advocacy count as advocacy

Definition	Agreement
European Parliament definition	63.6%
European Commission definition	65%
Facebook definition	62%
Twitter definition	66.7%

Table 2: Study 1: Agreement. The table shows percentages of ads workers agreed on in total for all three datasets

Definition	Political	Adver.influence	Mess.influence	Divisive
European Parliament definition	357	87.6%	82.1%	84.9%
European Commission definition	365	88.4%	82.3%	85.9%
Facebook definition	364	88.7%	83.8%	86.3%
Twitter definition	354	86.2%	80.7%	79.5%

Table 3: Study 2: Facebook dataset

Definition	Political	Adver.influence	Mess.influence	Divisive
European Parliament definition	43	58.9%	54%	57.6%
European Commission definition	59	87.5%	81%	79.7%
Facebook definition	65	92.9%6	88.9%	86.4%
Twitter definition	38	58.9%	54%	52.5%

Table 4: Study 2: IRA dataset

Definition	Operational†	Advocacy
European Parliament definition	4.8%	50.7%
European Commission definition	19%	70.4%
Facebook definition	9.5%	62%
Twitter definition	4.8%	52.1%

Table 5: Study 3: NGO dataset. Operational† means that an ad is only operational and does not have advocacy label.

[9] Ed Summers. 2018. *irads*. https://github.com/umd-mith/irads/blob/master/README.md

[10] Twitter. 2023. *Political content*. https://business.twitter.com/en/help/ads-policies/ads-content-policies/political-content.html

[4] House Intelligence Committee. 2018. *Exposing Russia's Effort to Sow Discord Online: The Internet Research Agency and Advertisements*. https://democrats-intelligence.house.gov/social-media-content/

[5] Facebook. 2022. *Ads about social issues, elections or politics*. https://www.facebook.com/business/help/167836590566506

[6] European Parliament. 2022. *Committee on Culture and Education*. https://www.europarl.europa.eu/doceo/document/CULT-AD-735573_EN.pdf

[7] Prolific. 2022. *Prolific*. https://www.prolific.co

[8] Qualitrix. 2022. *QualtricsXM*. https://www.qualtrics.com