

# Trade-offs in Automating Platform Regulatory Compliance By Algorithm: Evidence from the COVID-19 Pandemic

Cecere, G.\*      Jean, C.†      Lefrere, V.\*      Tucker, C.‡

## Abstract

Digital platforms have experienced pressure to restrict and regulate political ad content. In a static environment, using algorithms can help platforms more quickly and easily achieve regulatory compliance. However, in a dynamic context, the rigidity of complying with regulations by having to pre-specify the parameters that algorithms use as inputs, may pose challenges. We study the effect of algorithmic regulation of ad content in times of rapid change where digital ad venues need to identify sensitive ads that should be subject to more restrictive policies and practices. We collect data on European and American ads published in the Facebook Ad Library. Our results show that ads run by governmental organizations designed to inform the public about COVID-19 are more likely to be banned by Facebook’s algorithm than similar ads run by non-governmental organizations. We show that algorithmic (mis)classification of ads leads COVID-19-related ads to be disqualified. We suggest that algorithmic inflexibility towards categorization in periods of unpredictable shifts worsens the problems of trying to achieve regulatory compliance using algorithms.

**Keywords:** Algorithmic Decision-Making, Ad Ban, COVID-19

**JEL Codes:** M3, K2

---

\*Institut Mines-Telecom, Business School, 9 rue Charles Fourier, 91000 Evry, France. grazia.cecere@imt-bs.eu; vincent.lefrere@imt-bs.eu

†Grenoble Ecole de Management, 12 rue Pierre Semard, 38000 Grenoble, France. clara.jean@grenoble-em.com

‡MIT Sloan School of Management and NBER, 77 Massachusetts Ave, Cambridge, MA 02139, United-States. Please see <https://mitgmtfaculty.mit.edu/cetucker/disclosure/>. cetucker@mit.edu

# 1 Introduction

Advertising-supported digital platforms are under increased scrutiny when it comes to political advertising. As a result, various digital platforms have taken extreme steps to reassure authorities and users that political ads will be appropriately restricted and monitored. To ensure this monitoring, platforms rely on algorithms. Algorithms are increasingly used for prediction and classification, and for these purposes have the potential to match or exceed human performance (Brynjolfsson *et al.*, 2021). Though often research has focused on the advantages of algorithmic accuracy relative to human decision-making (Kleinberg *et al.*, 2018), another advantage of the use of algorithms is that they are inherently scalable (Berente *et al.*, 2021). This scalability is attractive to digital platforms trying to achieve regulatory compliance. Digital platforms often find regulatory compliance challenging, due to the sheer number, velocity and scale of interactions (Rao and Reiley, 2012). Many digital platforms establish their own rules to regulate sensitive advertising based on existing traditional media-based advertising regulations. Misleading information related to sensitive issues such as health issues can challenge information security Petratos (2021) thus it is important to investigate how this type of sensitive information is managed by algorithmic decision-making of large digital platforms.

However, in dynamic contexts, algorithms may be less effective at successful regulation, either because the inputs they were trained on are no longer pertinent, or because the outputs they produce are no longer appropriate (Agrawal *et al.*, 2018). We investigate the potential consequences of this for the use of algorithms to achieve regulatory compliance on digital platforms. We study the specific context of advertising. Using algorithms for regulatory compliance in advertising markets has become significant, especially given recent controversies over the interaction between politics and advertising on digital platforms (Isaac, 2019). The response to these controversies show the difficulty of regulating political ad on platforms. For example, Twitter has stopped accepting advertising with political content<sup>1</sup> and Facebook and Snapchat have implemented policies requiring la-

---

<sup>1</sup><https://www.eipartnership.net/policy-analysis/evaluating-transparency-in-platform-political-advertising-policies>, November 11, 2021.

belonging indicating the origin of all political ads.<sup>2</sup> However, to curate and regulate ad content deemed of potential political significance, a digital platform must first identify algorithmically which ads are subject to political regulation. Platforms face the further challenge of regulating themselves preemptively by trying to follow regulations designed for an analog era. The scale and diversity of ads in the digital era is far greater than in the analog environment in which political regulatory policies were developed.

We study use of algorithms to regulate sensitive ad content on Facebook. We are interested in the unexpected effects of algorithmic ad classification on ad display. Facebook uses an automated ad control procedure which classifies the ads and decides whether the ad content falls into the “*Social Issues, Elections or Politics*” (“SIEP”) category. If the algorithm classifies the ad into the SIEP category, or if the advertiser self-declares that the ad belongs to this category, then the ad is subject to the political ads policy. Facebook states that only certain COVID-19-related ads fall in the SIEP category and thus require a disclaimer. Therefore, not all COVID-19-related ads should be subject to the political ads policy. COVID-19-related ads require a disclaimer if they “*include political or electoral content, or if the content qualifies as a social issue ad.*” Advertisers are required to disclose who paid for the ad with a “*Paid for by*” label. If advertisers fail to comply with these requirements, the ad is disqualified.

We investigate the performance of this automation of political ad regulation in the context of a large health-related systemic shock, the COVID-19 pandemic. We examine 1,145,072 ads recorded in the Facebook Ad Library from January to June 2020 that were either related to the COVID-19, or were not related to the pandemic but were run by an organization which had published at least two ads related to the pandemic. We show that in the case of COVID-19-related ads, those posted by governmental organizations are more likely to be disqualified than those posted by non-governmental organizations. We conduct a text analysis which shows that the disqualified ads are likely to contain information related to COVID-19. Additionally, we use human annotation to show that

---

<sup>2</sup><https://businesshelp.snapchat.com/en-US/article/political-issue-ad-guidelines>, <https://www.facebook.com/business/news/requiring-authorization-and-labeling-for-ads-with-political-content>, May 16, 2020.

the disqualified ads were unlikely to have contained misleading information, so the disqualifications seem to not be driven by the advertiser’s likely intention. We show that disqualification has important implications as advertisers are unlikely to resubmit ads, and resubmitted ads are likely to remain disqualified. We augment our data using human annotation by MTurk workers. We identify a random subsample of COVID-19 ads in the SIEP category to check whether an ad is categorized as SIEP-related. Human annotation of ads reveals that the algorithm misclassifies 12% of the random subsample of ads as being SIEP-related. The algorithm’s misclassifications reduce display of important health information, with important social and health consequences for the population (Breza *et al.*, 2021).

These results highlight the difficulties involved in algorithmic regulation of potentially sensitive ad content in a global context. Algorithmic control of content may not adapt to new content in the context of a shock. Our findings have three main implications for marketing literature. First, the use of algorithms has unintended consequences on the type of ads banned, preventing governmental organization from promoting messages that are important and beneficial to public health and safety. Second, it raises questions about whether analog-era regulations are appropriate for algorithmic evaluation of online advertising, especially in relation to public health. Third, the classification of ad content is worsened in a dynamic context leading marketers and platform owners to be more careful about how algorithms curate information in a context of a shock.

Our paper is informed by, and builds upon, three streams of marketing and management literature. Early work such as Nelson (2003); Frank (2008); Goldfarb and Tucker (2011) and later work such as Chiou and Tucker (2018); Tuchman (2019) have studied the effect of ad bans initiated by governmental organizations to try to regulate the advertising of products that might harm consumer welfare, such as alcohol and tobacco ads. In contrast, we are studying the effects of a platform trying to implement digitally, analog-era rules that were originally designed to regulate political ads in an offline setting. Rather than looking at the effects of a ban, this study is to our knowledge the first to investigate how the use of algorithms to achieve regulatory compliance through ad ban can be ineffective

in a dynamic environment.

Second, the present paper builds also on a literature on algorithmic and improved platform transparency. Understanding algorithmic decision-making is a key issue in the online digital market (Choi *et al.*, 2020; Ma and Sun, 2020; Zhang *et al.*, 2021; Liu *et al.*, 2021). Previous work provides evidence of the existence of uneven distribution of ads based on ethnic origin (Sweeney, 2013) and gender (Datta *et al.*, 2015), although to interpret this as necessarily intentional discrimination and bias would be inaccurate (Lambrecht and Tucker, 2019; Cowgill and Tucker, 2019). The use of algorithms for classification, and potential distortions therein can induce unintended effects in terms of the type of information distributed to individuals (Berman and Katona, 2020; Kopalle *et al.*, 2022). While algorithm-based prediction is an important issue which has received much research attention (Fang and Hu, 2018; Hansen *et al.*, 2021; Zhang *et al.*, 2021), the present article tackles the role of algorithms in information (mis)classification. To our knowledge, the present paper is the first to study the role of algorithmic decision-making in relation to “SIEP” ad content. We contribute to this body of work by highlighting the risks of data misclassification by asking how algorithmic evaluation of ad content can have unintended consequences on the type of ad banned.

The third stream of literature we build on and are informed by, focuses on platforms and content regulation. Platforms have an underlying coring need to use technologies to make interactions go well (Gawer and Cusumano, 2015; Tucker, 2020), and, as platforms’ use of data increases, it is hard to scale such coring activities (Agarwal and Dhar, 2014). Not only do such algorithms help platforms scale, but they may help algorithms prevent bias in interactions (Fu *et al.*, 2021). However, beyond the business imperative to make interactions on platforms go well, platforms also have to comply with regulations. For example, political advertising is regulated in the US. There has been less research on the question of how algorithms fare in implementing regulatory compliance. De Vaujany *et al.* (2018) articulate the need for more research on the operation of the regulatory process as it becomes increasingly computer-mediated, but there has been little research on how algorithms actually perform at the task of regulatory compliance. We therefore

contribute to this literature by studying the use of algorithms to implement compliance with existing platforms rules at scale.

Fourth, this paper contributes to the economics of information security. Algorithmic mechanism design is a critical challenges for platforms (Anderson and Moore, 2007) as they need to ensure that no misalignment exists between the way the algorithm was designed to work and the final expected outcome. This has important security concerns when the platforms engage in self-regulation where adverse selection can have undesirable effects. As an example, Edelman (2009) shows in the context of a self-regulatory initiative that websites that seek and obtain trusted certifications are in fact significantly less trustworthy than those that waive certification. These effects have externalities in terms of information asymmetries which plays a large role in information security problems (Moore, 2005). We contribute to this literature by showing how algorithmic decision-making, in a context of platform self-regulation, can lead to a large number of false positives where trustworthy information is less likely to appeared on the platform. This has externalities on the platform user who benefits from a lower level of information quality in a crisis context where access to sensitive information is essential.

Our paper is organized as follows. Section 2 presents the institutional background. Section 3 describes the data collected via the Facebook Ad Library API. Section 4 presents the results of the data analysis. Section 5 study the potential mechanisms explaining our results. Section 6 concludes.

## **2 Institutional Background**

Facebook has evolved rules to restrict certain types of ad content, and in particular has developed rules for ad content it considers to be of national significance, whether the ad is posted on behalf of a political candidate or when the ad content is related to health or social issues. The breadth of this categorization can be seen in that Facebook applies its rules to all ad content which “*relates to any national legislative issue of public importance in any place where the ad is being run.*” This echoes language used in traditional US regulation of political ad on television channels (Oxenford, 2019), including 47 USC s.

315, “Candidates For Public Office,” which requires the licensee to maintain and make available for public inspection, a complete record of requests to purchase broadcast time which “*communicates a message relating to any political matter of national importance.*” Each traditional political ad must contain the disclaimer.<sup>3</sup> This can be done manually in an analog context format where the high costs of TV or radio ads limit the number and variety of ads. However, in a digital context it is costly for platforms to review all ads manually. In 2020, Facebook had 9 million active advertisers on its platform.<sup>4</sup> Ads on digital platforms tend to be more varied and numerous simply because the cost of targeting and distribution is much lower in the digital space which requires platforms to use algorithms to automate content evaluation.

Although as of yet there has been little explicit digital political ad regulation in the US, the Federal Election Commission (FEC) is undertaking different initiatives to regulate political ad spending on the Internet.<sup>5</sup> To increase ad transparency, Facebook, Google and Twitter choose to preemptively adopt their own rules based on political ad regulation which applies to traditional media such as television, radio and newspapers. Facebook requires advertisers to indicate a disclaimer in ads related to matter of national significance. Facebook also stores all ads in a public library.

Facebook advertisers with SIEP-related ads are required to undertake an ad authorization process which has to indicate their identity, location, and Facebook page address, in order to have a “*Paid for by*” label that can be attached to a given ad. The platform verifies the identity of the advertiser and authorizes them to run an ad with matter of national significance.<sup>6</sup> This authorization process for ads in this category is based on the traditional political ad rules. The disclaimer must identify in a transparent and easily understandable format the organization paying for the ad and should not include URLs

---

<sup>3</sup><https://www.fec.gov/help-candidates-and-committees/advertising-and-disclaimers/>, November 21, 2021.

<sup>4</sup><https://www.statista.com/statistics/778191/active-facebook-advertisers/>, November 21, 2020.

<sup>5</sup><https://www.ncsl.org/research/elections-and-campaigns/digital-political-ads.aspx>, December 10, 2021.

<sup>6</sup><https://www.facebook.com/business/m/one-sheeters/ads-with-political-content-EU>, May 2, 2020.

or acronyms, offensive language, or opaque words or phrases.

While there may be differences in application between countries or even differences in local law regarding the regulation of online ad content, this framework is general and applies to all ads, regardless of advertiser’s country or the targeted country.

## 2.1 Algorithmic Determination of National Significance

Before publication, all paid ads are reviewed by an automated ad-screening system to verify their compliance with Facebook’s general advertising policy, followed by algorithmic judgment about whether the ad falls into a special ad category. Ads containing information of national significance are subjected to an additional review process to verify compliance with the rules governing political ads on traditional media.<sup>7</sup> SIEP content covers a broad category of ads and includes social issues ads defined as paid content which “seek[s] to influence public opinion by including ad content that discusses, debates or advocates for or against a social issue”.<sup>8</sup> These rules were applied to COVID-19-related ads since the virus emerged and began to spread.<sup>9</sup>

Before launching an ad campaign, advertisers self-determine the inclusion of their ad in a particular category. To indicate content related to SIEP, the advertiser checks the box “*This ad is about social issues, elections and politics*”. Figure 1 reflects Facebook general guidelines and shows the checkbox.<sup>10</sup>

---

<sup>7</sup><https://medium.com/centerforcooperativemedia/what-small-publishers-need-to-know-about-facebooks-policy-on-ads-with-political-content-180874bf63c3>, May 2, 2020.

<sup>8</sup>[https://www.facebook.com/business/help/213593616543953?locale=en\\_GB](https://www.facebook.com/business/help/213593616543953?locale=en_GB), October 31, 2020.

<sup>9</sup>These rules were slightly adapted during the pandemic. <https://www.advertisemint.com/the-list-of-covid-19-ads-and-posts-banned-by-facebook/>, October 31, 2020.

<sup>10</sup>This checkbox was the checkbox available to advertisers on the Facebook Ads Manager at the time of the data collection.

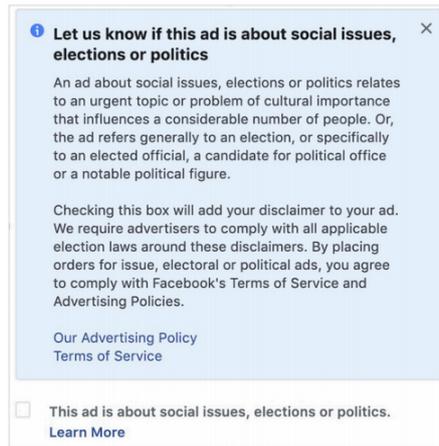


Figure 1: Ads Related to SIEP: Checkbox from Facebook

If this box is not checked because the advertiser does not consider that the ad is related to social issues, elections or politics but the automated ad-screening system classifies it as to be political or related to a social issue, the ad is placed in the SIEP category. Finally, the algorithm checks for the inclusion of an appropriate disclaimer. Figure 2 summarizes the algorithm's role in ad control.

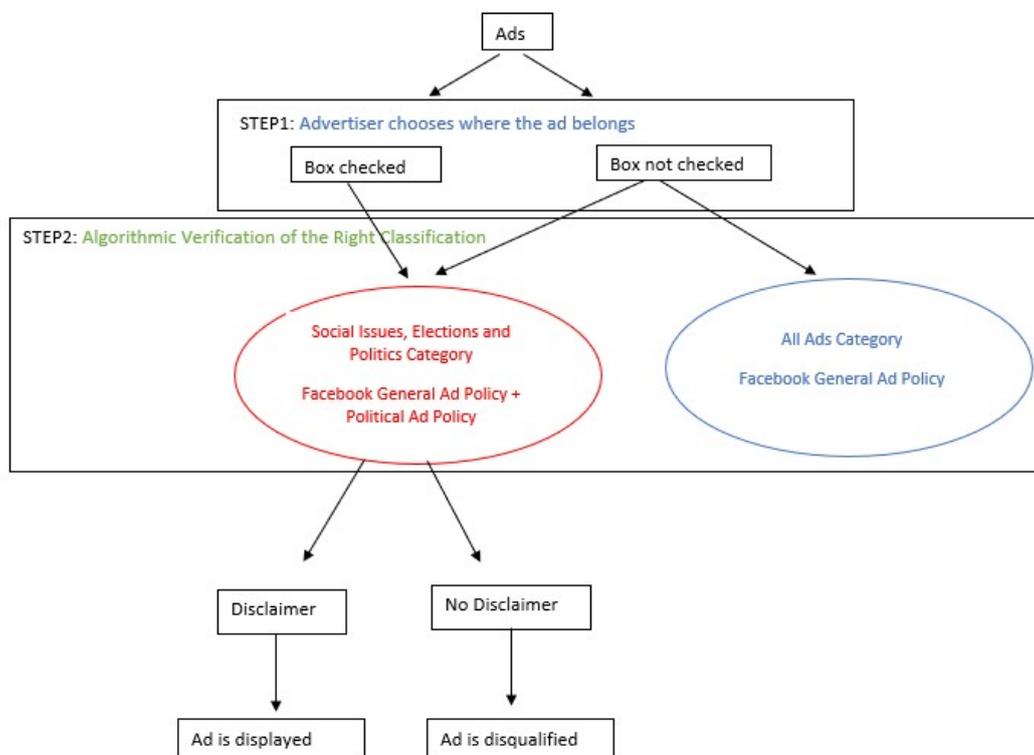


Figure 2: Algorithmic Decision-Making Process for Ads

If the ad is identified as political or relating to a social issue but does not include a disclaimer, it is rejected and labeled “*This ad ran without a disclaimer*” as illustrated in Figure 3. In this case, the advertiser receives notification of the ban on the grounds of non-compliance with the SIEP category rules. As previously mentioned Le Pochat *et al.* (2022) highlight that algorithmic classification of ads can lead to *false positives* with ads wrongly included in the SIEP category. In particular, they estimate that in the U.S., 55% of detected ads were incorrectly marked as political (false positives) and that Facebook misses 116,963 ads from clearly political advertisers (false negatives), with considerably worse performance outside the U.S.

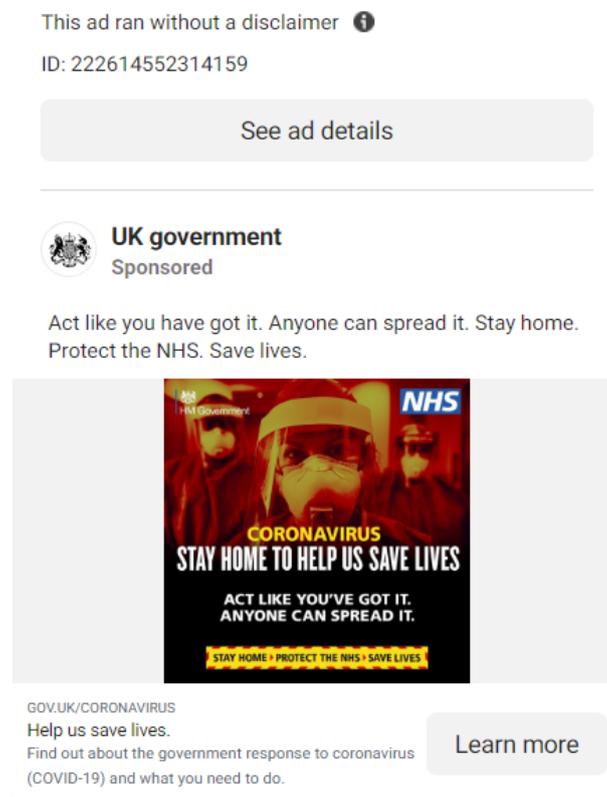


Figure 3: COVID-19-related Ad Ran Without a Disclaimer

Even before the COVID-19 pandemic, Facebook’s political ad policy was considered broad and included health content. For example, a health center was initially blocked for running ads without a disclaimer to raise awareness of PrEP, an FDA-approved anti-HIV medication sold under the brand name Truvada.<sup>11</sup>

<sup>11</sup><https://www.theguardian.com/technology/2019/oct/31/facebook-prep-ads-instagram-political>, May 16, 2020.

## 2.2 COVID-19 Ads

Facebook applies political ad regulation to certain ads to ensure advertiser transparency about ad origin. COVID-19-related ads are included in this category if the algorithm judges them to include “*political or electoral content, or if the content qualifies as a social issue ad*”. In our empirical setting, we can measure whether SIEP-related ads were banned by checking for “*Paid for by*” information. We identify the type of entities that posted all ads. This includes public authorities such as national ministries of health which used social media platforms to diffuse COVID-19-related information to inform citizens about the dangers related to COVID-19 and advise about how to avoid infection. Governmental organizations are increasingly posting ads to communicate with individuals.<sup>12</sup> The European Parliament was the governmental body that spent the most on Facebook ads in France, Italy and Germany between March 2019 and December 2020.<sup>13</sup> Therefore, it is useful to understand how algorithms apply political (analog) ad regulation designed originally to regulate political ads in the context of promoting health-related content on social media platforms such as content related to the spread of COVID-19.

## 3 Data

We gathered data using the Facebook Ad Library API which provides access to all SIEP-related ads (see Figure 11 in Appendix A). To identify ads related to COVID-19, we searched on the keywords “*coronavirus*” and “*covid*”<sup>14</sup> which were the most frequently searched terms on Google Trends in March 2020. We collected ads posted in Europe and in the US from January (when the first COVID-19-related ads were published) through June 2020. We decide to study the early spread of infections to assess how the algorithm reacted to the emergency.<sup>15</sup> In early February 2020, the US declared the coronavirus outbreak a public health emergency. In March 2020, the World Health Organization declared the COVID-19 outbreak a pandemic and many countries such as Italy, France,

---

<sup>12</sup><https://blog.hootsuite.com/social-media-government/>, December 26, 2020.

<sup>13</sup>In our sample the European Parliament spent on average €1,394 on COVID-19-related ad campaigns (total budget of €581,218) and they spent €1,601 on ad campaigns not related to COVID-19 (total budget of €674,758).

<sup>14</sup>These two words do not need to be translated.

<sup>15</sup>We conducted our empirical analysis excluding January and the results still holds.

Germany, and the United Kingdom announced COVID-19 lockdowns.

Based on this initial sample, we collect a unique identifier for each advertiser. We consider only those advertisers that ran more than one ad related to COVID-19 during the period January-June 2020.<sup>16</sup> This empirical strategy allows us to compare advertisers who were “*equally*” active in extensive outreach.

We identify Facebook pages for 10,825 out of the 11,122 advertisers. We retrieved all ads posted by a given advertiser, including ads unrelated to the pandemic which we used as the reference group. To collect more information on the advertiser running the ad, we retrieved advertisers’ Facebook pages to identify the entity’s category. In 2013, Facebook introduced the “*verified badge*” (“*blue check*”) for advertiser pages. This certifies accounts as belonging to the public figure, celebrity, global brand, or institution.<sup>17</sup>

Our final sample includes 1,145,072 ads posted on Instagram and Facebook, across Europe and the US. We collected detailed characteristics of each advertiser and the ads posted. The sample includes both active and inactive ads. Table 1 provides summary statistics. We split the summary statistics by whether or not advertiser is a governmental organization.

*Disqualified ads* takes the value 1 if a given ad does not have a disclaimer and 0 otherwise. *COVID-19 ad* takes the value 1 if the ad contains either of the terms “*covid*” or “*coronavirus*.” We identify governmental organization using the categories indicated on the advertisers’ Facebook page and include the classes “*Public & Government Service*” or “*Governmental Organization*”. *Governmental Organization* takes the value 1 if the Facebook page belongs to a governmental organization. We do not include in this category politicians or political parties.

To determine whether the advertiser page is official, we use the variable *Verified badge*,

---

<sup>16</sup>This threshold corresponds to the 5th percentile distribution lower bound which corresponds to 8,822 advertiser pages. We exclude SmartNews the largest advertiser in our initial sample with 233,855 ads, since it used an excessive number of variations of wording to localize each of its ads. We provide in section E the main results without this cutoff.

<sup>17</sup>This verification process is different from running ads in the SIEP category. While the verified badge allows to trace and verify who is behind a Facebook page, the “Pay for by” in the SIEP category is a requirements for posting ads on a political matter.

it takes the value of 1 if the ad is paid by an organization with the verified “*blue check*” mark. To identify whether the ad was published in Europe or in the US, we create the dummy variable *Europe*, they account for 23.66% of the overall sample. If the ad run in both the Europe and in the US, we attribute it to the US.

Table 1 shows that 1.4% of ads are disqualified overall but this percentage is of 4% when it comes to governmental organizations. A higher share of ads ran by governmental organizations are COVID-19-related compared to the subsample of other advertisers. While the disqualification rate is higher for governmental organizations, the large majority of governmental organizations have a verified badge.

Table 1: Summary Statistics

Variable	Overall		Non-Gov. Orga		Gov. Orga		Min.	Max.
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.		
Disqualified ads	0.014	(0.119)	0.014	(0.117)	0.04	(0.195)	0	1
COVID-19 ad	0.164	(0.37)	0.164	(0.37)	0.222	(0.416)	0	1
Gov. Orga	0.011	(0.106)	-	-	-	-	0	1
Verified badge	0.535	(0.499)	0.533	(0.499)	0.731	(0.443)	0	1
Europe	0.237	(0.425)	0.235	(0.424)	0.413	(0.492)	0	1
Observations	1,145,072		1,131,938		13,134			

As we restrict our sample to advertisers displaying at least two COVID-19-related ads, we do not include the more general pool of advertisers. This empirical strategy results in conservative measures, since we exclude advertisers who are more familiar with general product ads and less familiar with health-related ads. We believe that our sample provides at worst lower-bound results. Advertisers who choose to advertise COVID-19-related ads are more likely to be familiar with health and social issues, and therefore are more likely to be aware of Facebook ad regulations. This suggests that ads run by governmental organizations should be less likely to be banned. While the overall percentage of disqualified ads is low, it should be noted that a quarter of the advertisers in our sample are disqualified at least once. Also, in our sample, governmental organizations are largely represented among verified advertisers. That is, even among the advertisers most likely to be informed about Facebook’s specific rules for this type of content, a significant

number of advertisers are disqualified.<sup>18</sup>

## 4 Results

Figure 4 depicts our main result. The y-axis shows the proportion of disqualified ads in our sample. It shows the proportion of disqualified ads according to whether they are related to the COVID-19 pandemic - an event that was unexpected and whose categorization as a political issue or not was somewhat ambiguous at the time of the initial outbreak of the virus. Overall, COVID-19-related ads were more likely to be disqualified.

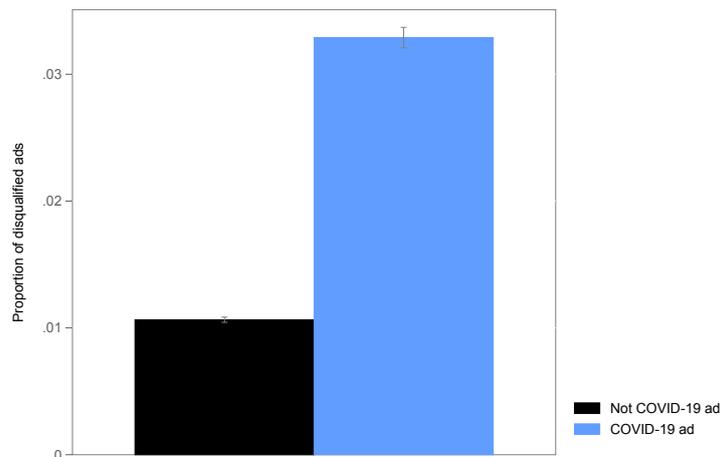


Figure 4: Disqualified Ads

Figure 5 shows the proportions of disqualified ads run by governmental and non-governmental organizations. Ads posted by governmental organizations were more likely to be disqualified than ads placed by non-governmental organizations. This pattern of COVID-19-related ads being disqualified is even greater if it was a governmental organization that was placing the ad. This is surprising as governmental organizations are often law-making bodies and might be expected to be more compliant with social media advertising rules. These results show that in period of rapid change when there is a lack of consensus about the definition of sensitive ads that should be subject to earlier policies, an algorithmic definition of sensitive content may be problematic. It's worth noting that business companies are not particularly inclined to place ads in this category. Taking Coca-Cola as an example, we observe only 1 ad in the SIEP category since 2018. This may be explained

---

<sup>18</sup>In 2022, only 3.25% of disqualified advertisers were no longer active on the platform.

by the fact that this type of advertisers do not necessarily run ads about a general matter which may fall in the SIEP category.

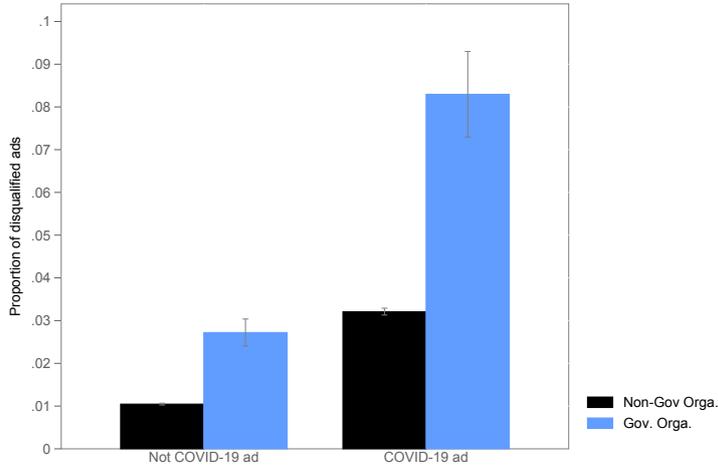


Figure 5: Disqualified Ads by Governmental Organizations

These results are robust to a broader set of non-English keywords available in Table 8 (see Figures 15 and 16) in Section D.

#### 4.1 Does the Ad Disqualification Reflect a Lack of Knowledge about Facebook Policy?

Our results suggest that COVID-19-related ads published by governmental organizations are more likely to be disqualified. To assess whether this in turn reflects a lack of knowledge about Facebook’s SIEP advertising policy, we look at whether the organization attempted authentication before placing the ad. We measured the presence of a verified badge which would confirm authentication by Facebook of an account of a public figure, media company, or brand. We use this data to measure whether organizations less knowledgeable about Facebook rules were more or less likely to omit a disclaimer.

Among the subsample of advertisers with a verified badge, Figure 6 shows that the proportion of disqualified ads is higher if the ad was placed by a governmental organization, and especially if it was COVID-19-related. It should be noted that among ads run by governmental organizations, 73.11% of ads are run by a verified advertiser. This suggests that the large proportion of disqualified ads cannot be explained by the fact that the

governmental organization was not known to Facebook, or was unable to follow the verified account procedure. We suggest rather that the algorithm might apply an automatic ad classification based on COVID-19-related keywords rather than complex algorithmic classification.

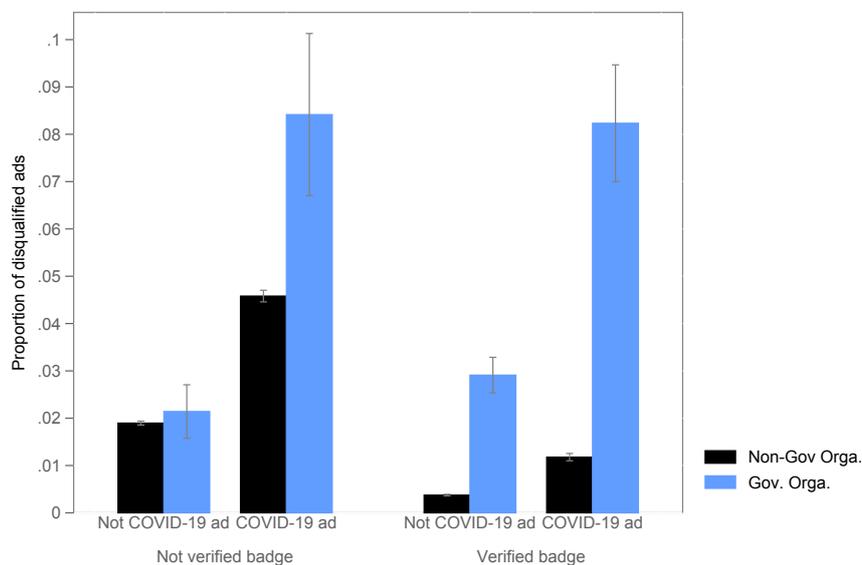


Figure 6: Disqualified Ads by Verified Advertisers

To further explore the mechanism leading to this result, we explore whether the proportions of COVID-19-related disqualified ads are the same in Europe and in the US. Figure 7 shows that COVID-19-related ads run by governmental organizations based in Europe were more likely to be disqualified than ads run by organizations in the US. One interpretation is that European governmental organizations are less familiar with Facebook’s political ad policy, echoes language used in US traditional regulation of political advertising, and therefore may not have realized that their ad would be judged by the algorithm as subject to regulation. In novel and dynamic contexts such as the COVID-19, the lack of available knowledge and experience to train algorithms could lead to excessive numbers of banned ads, especially in Europe.

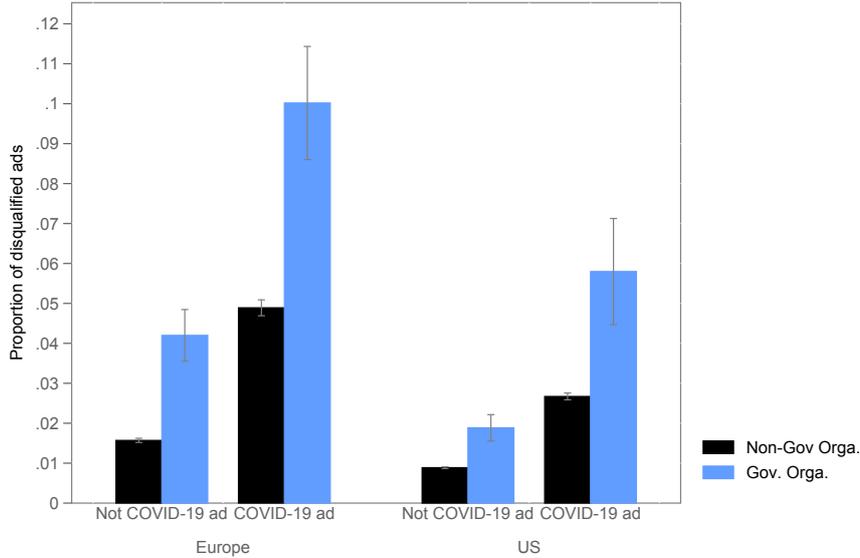


Figure 7: Disqualified Ads by European and US Advertisers

## 4.2 Logit Estimates

To study the relationship between COVID-19-related ads and the probability of an ad to be disqualified, we use our cross-sectional data to estimate a logistic model. Table 9 presents the results of the marginal effects. Column (1) shows that COVID-19-related ads are more likely to be disqualified. This finding aligns with the earlier graphs we presented. The likelihood of being disqualified also increases when ads are run by a governmental organization. Ads posted by advertisers based in Europe are more likely to be disqualified. Then, we estimate the probability model of disqualified ads separately for ads posted by governmental organizations and non-governmental organizations. In columns (2)–(3), we show that the results about COVID-19-related ads and European advertisers remain robust when we split the estimates by whether the advertiser is a governmental organization or not. The estimates suggest that ads posted by European governmental organizations are more likely to be disqualified where the variable *Europe* has larger marginal effect for the subsample of governmental organizations.

Table 2: Ads Run by Governmental Organizations Are More Likely to Be Disqualified

	Overall	Gov. Orga	Non-Gov. Orga
	(1)	(2)	(3)
COVID-19 ad	0.015*** (0.001)	0.037*** (0.004)	0.015*** (0.000)
Gov. Orga.	0.012** (0.006)		
Europe	0.008*** (0.002)	0.023*** (0.004)	0.008*** (0.000)
Week Fixed Effects	Yes	Yes	Yes
Log-Likelihood	-80509.557	-1881.946	-78455.166
Wald chi-squared test	500.14	728.03	11839.30
Observations	1,145,072	12,979	1,131,938

*Notes:* Logit Estimates reported as marginal effects calculated at mean. Dependent variable is whether the ad is disqualified. Column (1) reports the robust standard errors clustered at the advertiser level. Column (2) omits 155 observations as the last week of January 2020 does not include any COVID-19 ads run by governmental organizations. Column (2) and (3) report robust standard errors. Significance levels: \*\*  $p < .05$ , \*\*\*  $p < .01$ .

## 5 Underlying Mechanisms

When algorithmic-decision making is used in a highly dynamic environment, there is a potential of creating harm by preventing the distribution of socially beneficial ads, it is important to consider and rule out alternative explanations for the phenomenon observed in the data. First, we investigate whether disqualified ads contain potentially misleading claims. Second, we check whether the algorithm is likely to disqualify ads served in smaller ad campaigns. Third, we check whether our results are driven by advertisers that run multiple versions of very similar ads. Four, we perform text analysis to check whether the content of ads from government organizations is different from ads ran by other advertisers. Five, we check whether disqualified ads are likely to resubmitted. Finally, we check algorithmic (mis)classification in the SIEP category.

### 5.1 Were the Banned Ads Potentially Misleading?

We show that COVID-19-related ads run by governmental organizations were more likely to be banned. This suggests that ads are banned not because of undesirable content. It is likely that ads submitted by governmental organizations are intended to inform

the population. However, to check whether this is an optimistic view of the role of governmental organizations, we check whether these ads included potentially misleading claims.

We randomly selected 1,000 disqualified COVID-19-related ads from our sample. We manually checked whether these ads include any misleading claims. We follow the guidelines provided by the statement of the Australian Competition and Consumer Commission (ACCC) to distinguish between false and misleading claims.<sup>19</sup> We also rely on the FTC statement related to unproven COVID-19 claims which state that “*anyone who makes deceptive claims about the treatment, cure, prevention, or mitigation of COVID-19 is subject to civil penalties*”.<sup>20</sup> Manual checking revealed that only 5% of ads in this subsample contained potentially misleading claims, and all were published by non-governmental organizations.

This suggests that the majority of disqualified ads did not include any misleading claims. Our analysis suggests that the disqualified COVID-19 ads were mostly likely to inform individuals about the pandemic. Therefore, it seems that automated ad control is likely to reduce the diffusion of messages aiming to inform individuals. These results underline the problems related to algorithmic monitoring and filtering of content in dynamic contexts.

## **5.2 Are Larger or Smaller Campaigns more Likely to Be Disqualified?**

Ads related to COVID-19 may convey important health information. Therefore, if advertisers running large ad campaigns are more likely to be disqualified, this could affect diffusion of information related to the pandemic. We investigate this using additional data available on the Facebook platform. Alongside characteristics such as content, day of creation and location of targeted users, the Facebook Ad Library provides detailed performance information for ads placed on Facebook and Instagram. Impressions are broken down by age group, gender, location and the amount of money spent on the ad and the platform on which it is posted are available.

---

<sup>19</sup>See the definition: <https://www.accc.gov.au/consumers/advertising-and-promotions/false-or-misleading-claims>, October 10, 2022.

<sup>20</sup><https://www.ftc.gov/business-guidance/blog/2021/04/advertisers-stop-unproven-covid-claims-or-face-penalties-under-new-law>, October 10, 2022.

Figure 8 and Figure 9 show respectively the average number of impressions and the average amount spent by advertiser splitting the sample between non-governmental and governmental organizations. We create the variable, *Ban*, which takes a value of one if the advertisers have at least one disqualified ad. Overall, any differences in size – in terms of impressions and amount spent – are not significant, suggesting that advertisers whose ads are banned are not different from those whose ads are never disqualified. However, governmental organizations are likely to run larger campaigns in terms of numbers of impressions and the amount of ad budget spent compared to non-governmental organizations. Additionally, we present evidence in Figure 17 in Appendix E that governmental organizations and non-governmental organizations were likely to experience the same disqualification rate in early pandemic phase.

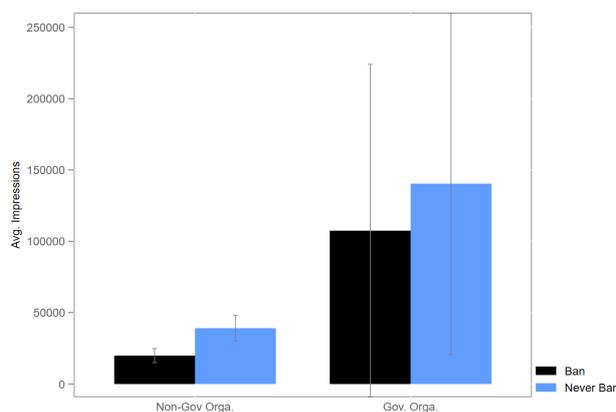


Figure 8: Average Number of Impressions

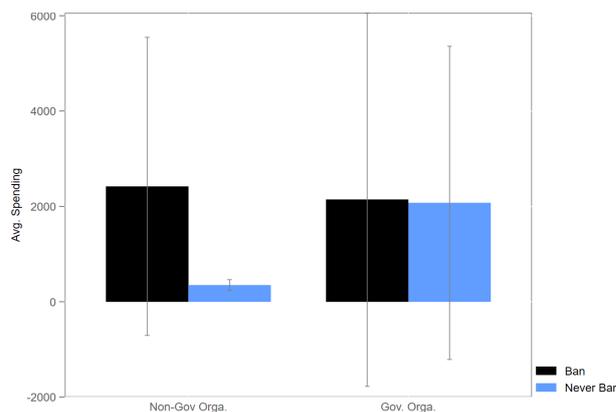


Figure 9: Average Spending

### 5.2.1 Examples of Governmental Organizations Banned or Never Banned

To provide evidence on the type of organizations that are disqualified, Table 3 shows the list of the largest organizations among governmental organizations in terms of amount spent. Governmental organizations are ordered in decreasing order divided in whether they have been disqualified at least once (*Ban*) or not.

As the pandemic continued, Facebook encouraged public organizations to use its platform for public service announcements (PSAs), and set up a process to enable free posts which contained this kind of information.<sup>21</sup>

<sup>21</sup><https://www.facebook.com/gpa/blog/tips-and-tools-for-public-service-announcements>, last accessed October 31, 2020.

Table 3: Largest Spending Organizations Among Governmental Organizations

Ban	Never Ban
US Centers for Disease Control and Prevention (CDC)	Kancelaria Premiera <sup>21</sup>
European Parliament	California Labor and Workforce Development Agency (LWDA)
UK Government	Európai Bizottság Magyarországi Képviselete <sup>22</sup>
Région Auvergne-Rhône-Alpes <sup>23</sup>	City of New York
Tobacco Free Florida	NYC Mayor’s Office

Notes: <sup>21</sup> Polish Chancellery of the Prime Minister  
<sup>22</sup> Hungary Representation for the European Commission  
<sup>23</sup> French region

This strategy aimed to allow a small number of trusted organizations such as CDC, UNICEF, WHO and national ministries of health to advertise COVID-19-related ads for free in order to prevent the spread of the virus and fight disinformation.<sup>24</sup> While CDC and national ministries of health are considered trusted organizations by Facebook, they have COVID-19-related disqualified ads. In particular, the Germany Ministry of Health (Bundesministerium für Gesundheit) has (among the governmental organizations) the largest number of disqualified ads with 90 disqualified COVID-19-related ads out of 163 COVID-19-related ads. Another example is the Luxembourg Ministry of Health, the algorithm disqualified all COVID-19-related ads.

### 5.3 Are The Results Driven By Ads with Duplicative Ad Content?

One concern about the interpretation of our results, is that many advertisers distribute multiple versions of very similar ads. They may do this because they are engaging in ‘a/b’ testing to understand whether certain types of ad content are more effective. Or advertisers may do this to make small tweaks such as personalizing ad copy to a particular state or geographic region. This might mean that advertisers who have the most versions of essentially the same ad, are both over-represented in our data, and also that these are the advertisers that might have the most expertise in designing ad copy. Table 4 shows governmental organizations by average amount spent based on whether or not they duplicated their ads. Column (1) shows governmental organizations that run a single version of their ad. Column (2) shows governmental organizations that reproduce a given ad multiple times. Governmental organizations such as the European Parliament,

<sup>24</sup><https://www.theverge.com/2020/3/4/21164496/facebook-coronavirus-world-health-organization-free-ads> and <https://about.fb.com/news/2020/12/coronavirus/>, October 31, 2020.

beside being among the organizations that have spent the most on the platform, are more likely to duplicate their ads and therefore exhibit expertise on how to use the Facebook ad platform. However, there is no significant difference in terms of disqualification rate between advertisers who duplicate or not their ads. Therefore, even advertisers who display technical know-how in terms of running multiple versions of an ad are not successful in avoiding disqualification.

Table 4: Duplication of Ads and Spending by Governmental Organizations

Single version (1)	Multiple versions (2)
Delaware Department of Health and Social Services	CDC
Folkhälsomyndigheten <sup>22</sup>	LWDA
Healthy Delaware	City of New York
Länsstyrelsen Östergötland <sup>23</sup>	European Parliament
Presidenza del Consiglio dei Ministri <sup>24</sup>	Európai Bizottság Magyarországi Képviselete
Trafikverket <sup>25</sup>	Kancelaria Premiera
Utah Department of Workforce Services	NYC Mayor’s Office
West Midlands Police	UK Government

*Notes:* <sup>25</sup> Swedish The Public Health Agency

<sup>26</sup> Swedish Regional Agency

<sup>27</sup> Italian Ministry Council

<sup>28</sup> Swedish Transport Administration

These multiple versions of ads may be problematic if they drive our main results due to their quantity. To check the robustness of our results, we focus on a subsample of ads which contain unique ad content, and do not have any close duplicates on a single day. The remaining subsample includes 406,740 unique ad contents. Figure 10 shows the proportion of disqualified ads in this subsample. Overall, 2.3% of unique ads are disqualified. The disqualification rate is about 4.45% for COVID-19-related ads. This percentage is equal to 6.29% for COVID-19-related ads run by governmental organizations.

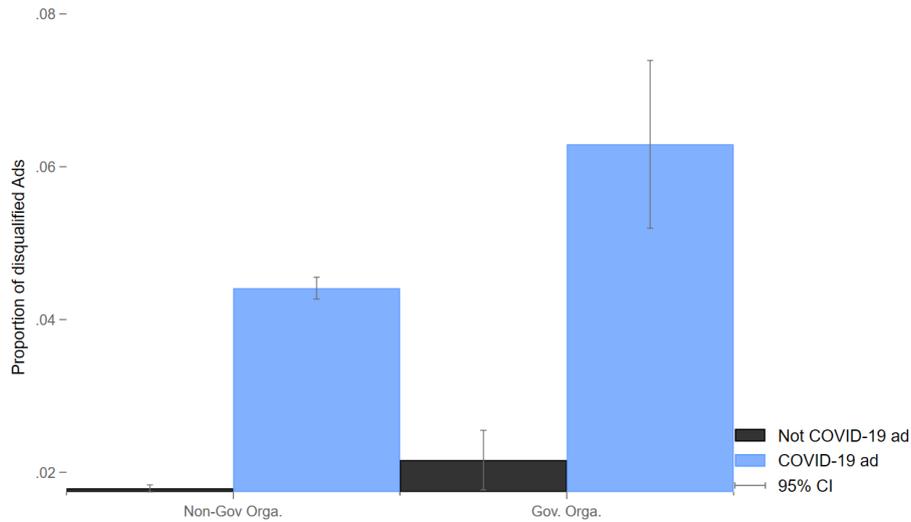


Figure 10: Disqualified Ads with Unique Ad Content

Figure 10 shows that the disqualification rate is higher for ads with non-duplicative content relative to earlier results. An explanation of this is that advertisers who automate their ad content are more likely to be experienced about the ways the platform applies its rules and therefore may understand how to comply with platform ad policy. Or alternatively, advertisers might initially test an ad, see if it is accepted, and then if it is accepted, duplicate its content in order to personalize it to, for example, different geographies. In general, it is reassuring that our results are not driven by a few pieces of ad content that were duplicated multiple times.

#### 5.4 Are Governmental Organization Ads Different from Other Advertisers?

One concern in our analysis is that ads run by governmental organizations are clearly different when they are related to COVID-19 compared to other ads which may explain why the rejection rate is higher. To address this concern, we check for differences between ads ran by governmental organization and non-governmental organizations. In our textual analysis we focus on English language ad content. English ad content accounts for 80.32% of the sample. We deleted stop words, removed all punctuation, and transformed upper case into lower case characters. We use GridSearch optimization which allowed us

to define the number of topics. We test between 2 and 15 topics, and to enable comparison we identify three topics (T=3) for each sub-analysis. Table 5 presents the top 10 words associated with the LDA analysis for COVID-19 disqualified ads submitted by governmental and non-governmental organizations. Among governmental organizations, we have three main topics, “*Coronavirus*”, “*Public authorities*” and “*Protection*”. We compare this result with COVID-19-related ads run by non-governmental organizations and we have created three main topics, namely “*News*”, “*Coronavirus*” and “*Business*”. Looking at the “*Coronavirus*” topic, we observe common words (indicated in bold in Table 5) used in the ads such as “coronavirus”, “health” and “people” and we fail to observe misleading keywords which could break compliance with Facebook guidelines for the subsample of governmental organizations.

Table 5: Top 10 Words from the Topic Analysis LDA-Rating Model

	Gov. Orga			Non-Gov. Orga		
	“ <i>Coronavirus</i> ”	“ <i>Public Authorities</i> ”	“ <i>Protection</i> ”	“ <i>News</i> ”	“ <i>Coronavirus</i> ”	“ <i>Business</i> ”
<b>coronavirus</b>	count		product	news	<b>coronavirus</b>	estate
<b>health</b>	determine		protection	county	<b>health</b>	market
state	service		<b>help</b>	break	<b>help</b>	create
county	county		order	stay	need	need
measure	water		cost	app	support	start
information	census		brand	million	time	time
<b>people</b>	committee		description	trust	<b>people</b>	<b>people</b>
emergency	emergency		abuse	install	business	investor
court	city		know	live	covid	invest
family	money		family	breaking	make	foreclosure

Table 6 presents a list of the top 10 words in our sample which is split between COVID-19-related and not COVID-19-related ads. We then split these subsamples into ads submitted by governmental organizations and non-governmental organizations and examined the most frequent words in the accepted and disqualified ads. Among COVID-19-related ads, we observe a lot of similarities in terms of word frequencies between ads that were or not disqualified run by governmental organizations. Comparing this word frequencies with non-governmental organizations, we observe a slight difference with “news” and “trump” which could be more sensitive to algorithmic ban while this is not what we

retrieve in our results.

Table 6: Most Frequently Used Words Used in Our Sample

COVID-19 ad				Not COVID-19 ad			
Gov. Orga.		Non-Gov. Orga.		Gov. Orga.		Non-Gov. Orga.	
Disqualified (1)	Not Disqualified (2)	Disqualified (3)	Not Disqualified (4)	Disqualified (5)	Not Disqualified (6)	Disqualified (7)	Not Disqualified (8)
covid19	covid19	covid19	covid19	count	take	break	make
coronavirus	coronavirus	coronavirus	coronavirus	determines	fill	latest	sign
pandemic	stay	estate	estate	important	health	stay	today
help	help	help	help	help	care	millions	help
protection	absentee	real	real	protection	get	breaking	president
health	health	investing	need	census	census	informed	need
available	together	people	people	water	counted	app	vote
family	ballot	market	us	county	us	county	us
county	information	news	trump	services	everyone	news	trump
public	safe	time	time	safe	nyc	local	time

## 5.5 Are Disqualified Ads Resubmitted and Displayed?

If an ad is rejected by the algorithm because the appropriate disclaimer was not incorporated, the advertiser is notified of this decision and can edit the ad by checking the disclaimer box and resubmitting the ad.<sup>29</sup> We study whether or not advertisers whose ads are rejected, are likely to correct their initial submission, or whether they are ultimately deterred from ever advertising.

We observe 9,460 ads with unique ad content that are disqualified. To understand whether ad disqualification has a significant effect in terms of the information not displayed to users, we check the percentage of ads resubmitted to the platform after an initial ban and the percentage of ads that were disqualified after resubmission. We count the number of times ads were resubmitted by advertisers. Table 7 details the descriptive statistics. Among this subsample, only 29.1% of disqualified ads are resubmitted at least once which means that 70.9% of ads are never resubmitted following an initial ban. Overall, advertisers resubmit an ad 3.1 times on average. In the case of COVID-19 disqualified ads, we find that 32.99% of them were resubmitted and advertisers realize 2.68 attempts on average. Among ads run by governmental organizations, 35.19% of ads were resubmitted with an average of 3.3 resubmissions.

<sup>29</sup><https://medium.com/fee-marketing-insights/facebooks-political-content-disclaimer-and-what-to-do-when-your-ads-get-rejected-8ac17396d03f>, March 18, 2021.

Table 7: Many Disqualified Ads Remained Disqualified after Resubmission

	Overall		COVID-19		Gov. Orga		N
	%	#Submission	%	#Submission	%	#Submission	
Resubmitted at least Once	<i>29.1</i>	3.1	<i>32.99</i>	2.68	<i>35.19</i>	3.3	9,460
Always Disqualified Ads	<i>87.7</i>	1.59	<i>92.58</i>	1.49	<i>78.97</i>	1.91	9,460
Ads Publish at the End	<i>12.3</i>	2.11	<i>7.42</i>	2.26	<i>21.03</i>	1.2	9,460

In this subsample, 87.7% of ads remains disqualified overall and were never displayed. This figure is of 92.58% for ads that are COVID-19-related and of 78.97% for ads run by governmental organizations. These results suggest, first, that COVID-19 ads are likely not to be displayed to users if they have been disqualified, and second, that governmental organizations are more likely than other advertisers to have their ads allowed on resubmission after an initial disqualification. On average, ads that are always disqualified were resubmitted 1.59 times while ads that were eventually run were resubmitted 2.11 times on average but only 1.24 times if they are run by a governmental organizations. COVID-19-related ads that are initially disqualified and then published are submitted to the platform 2.26 times.

This suggests that algorithmic rejection of ads has a systematic and prolonged effect on the type of ads shown on Facebook and Instagram. We have shown that this is not because governmental organizations do not understand the rules, since they are more likely than other advertisers to have their ads allowed following initial disqualification, but rather that those ads should never have been put into that category.

## 5.6 Algorithmic Classification Verification

First, we check whether COVID-19-related ads are not included in the SIEP ad category. Second, we rely on human annotation on Amazon Mechanical Turk to identify any algorithmic measurement errors.

### 5.6.1 Are COVID-19-Related Ads Frequently not Included in the “SIEP” Ad Category?

At the time of the data collection, the Facebook Ad Library gives access to US but not European data in the categories “*Housing*”, “*Employment*”, and “*Credit*”. In Europe, the

Facebook Ad Library has only two ad categories: “*All ads*” and SIEP ads.

We used this variation to see how often COVID-19-related ads appeared into different ad categories. We looked at ads in the US “*Housing*”, “*Employment*” and “*Credit*” categories using the keywords “*covid*” and “*coronavirus*” during the period January to June 2020. We identify 12 pandemic-related ads not included in the category “SIEP” ads. “*Housing*” accounted for 11 ads and “*Employment*” included only one ad. We do not find any ads related to COVID-19 in the category “*Credit*”. We also observe that the difference of COVID-19-related ads between categories “All ads” and SIEP-related ads is marginal and equal to 3,206 ads. This suggests that COVID-19 ads can be potential included in the “*All ads*” category thus do not require a disclaimer. However, most of the COVID-19 ads have fallen in the SIEP category.

### 5.6.2 Algorithmic Classification and MTurk Worker Annotation

Algorithmic classification of ads into different categories determines which policy will govern a particular ad. This raises questions about the accuracy of the algorithmic classification. To assess any algorithmic measurement error in classifying COVID-19-related ads as political, we rely on human annotation on Amazon Mechanical Turk.

We give MTurk workers Facebook’s definition of political ad for reference. Based on these Facebook guidelines, we asked one hundred MTurk workers to annotate each a random sample of 100 disqualified COVID-19-related ads and to indicate whether they considered the content to be political or not. We assigned the task to workers with an approval rate of above 99% (high reputation) in line with the suggestion in Peer *et al.* (2014). We also asked each worker to rate whether they were confident about their judgment. Figure 12 in Appendix B shows the instructions and Figure 13 shows how we ask MTurk workers to rate their confidence in their classification. We aggregate worker annotations and consider an ad as political if the majority of workers classified it as political. Among these 100 ads, 12% of ads are considered non-political by more than 50% of MTurk workers, and 93.6% of MTurk workers declare confidence in their response. Our results corroborate the findings in Le Pochat *et al.* (2022) which highlight that algorithmic classification of

ads can lead to *false positives* with ads wrongly included in the SIEP category.

We test MTurk workers’ abilities to understand Facebook guidelines and to properly classify ads. As before, we assigned the task to workers with an approval rate above 99%. We asked the MTurk workers to evaluate five ads that were clearly political ads (see Figure 14 in Appendix C). All ads are classified as political by the majority of MTurk workers, suggesting that MTurk workers are likely to understand Facebook rules and to correctly classify ads.

## 6 Discussion

We study the effect of algorithmic regulation of ad content in times of rapid change. Digital ad venues need to identify sensitive ads likely to be subject to more restrictive policies and practices. However, in periods of rapid change when there is a lack of consensus about which ads are sensitive and should be subject to previously drafted policies, using algorithms to identify sensitive content can be problematic. We suggest that algorithmic misclassification of ads leads COVID-19-related ads to be more disqualified during the pandemic because the algorithm judges that they were about an issue of national significance but did not include a disclaimer and thus does not comply with Facebook political ad regulation. This does not seem driven by the likely intentions of the advertiser. COVID-19-related ads posted by governmental organizations are more likely to be disqualified than those posted by non-governmental organizations. This is not because governmental organizations do not understand the rules but rather that given Facebook guidelines, advertisers do not believe that their ads fell into this category despite algorithmic classification of ad considers them as political. Additionally, the platform encourages the diffusion of information to reduce the spread of the virus and fight disinformation , or help “*to connect people to information from regional and local health organizations*”<sup>30</sup> by allowing free ads sponsored by World Health Organization and other national institutions. These results highlight the difficulties involved in algorithmic regulation of sensitive ad content in a global context. Our results not only show an impact on access to information

---

<sup>30</sup><https://about.fb.com/news/2020/12/coronavirus/#mark-post>, February 02, 2023.

because few ads are submitted and accepted by the platform after a disqualification, but they highlight that the existence of false negatives and false positives generated by the algorithm has a greater implication than the scope of our study regarding algorithmic classification. Algorithmic control of content might not adjust to new content in the context of a shock. This raises questions about whether analog-era regulations are appropriate for algorithmic evaluation of online advertising, especially in the area of public health. In terms of platform design, the platform needs to provide advertiser guidelines that are in line with what the algorithm can do, in order to avoid misclassification and disqualification of ads. The importance of these issues is heightened by our finding that the prohibition of certain ads can be significant in terms of the managerial and policy implications and the fact that a banned ad is unlikely to be resubmitted and likely to remain disqualified if the ad is resubmitted. This suggests that well-intentioned attempts to achieve regulatory compliance of ads via algorithm that include content of public importance can have unintended consequences, due to algorithmic misclassification of ads, especially in unstable and uncertain periods.

Our article adds to our understanding of the negative implications of the use of algorithmic decision-making in a dynamic environment. The need to understand how algorithmic decision-making works in a crisis becomes even more urgent if the information being managed is related to matters of national significance. Regulations based on automated decision-making may not be sufficient to allow appropriate judgments about banning ads during the COVID-19 pandemic or some other major exogenous shock. Overall, more interaction with humans is needed to complement automated decision-making (Cowgill and Tucker, 2017) to enable more trustworthy automated systems (Cowgill and Tucker, 2019), when trying to achieve regulatory compliance at scale – and particularly in dynamic contexts.

Our study has three main practical implications. First, our findings provide evidence that algorithms are unable to make appropriate judgments in fast changing environments such as the COVID-19 pandemic. Second, we show that automated algorithmic content filtering has drawbacks when trying to categorize content. This raises questions about

standards for and regulation of health-related content on digital platforms and the extent to which it should be part of what is considered political because it is an area of policy controversy in the US. Third, our results highlight that in times of crises there may be more need for intervention to manage algorithms.

Our study has also several implications for policy. Lefouili and Madio (2022) emphasize that policy makers should take account of certain economic trade-offs when designing a platform liability regime. In the US, the FEC has not yet defined clear rules to regulate online political ad content, therefore digital platforms have created their own rules based on traditional media regulation (Brennen and Perault, 2021). The EU Digital Services Act requires online platforms to adopt higher standards of transparency and accountability in relation to moderation of content and advertising, and algorithmic processes (Cabral *et al.*, 2021). We contribute to this debate by showing that algorithms trained on regulations designed for the analog era will be unable to correctly classify ad content on digital platforms, and that this results in unintended consequences for the type of content banned. We show that automated algorithmic content filtering has drawbacks when trying to classify content, especially in a dynamic environment or fast changing environments such as the COVID-19 pandemic. This highlights that in times of crises there may be more need for intervention to manage algorithms.

## Bibliography

- Agarwal, R. and Dhar, V. (2014). Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. *Information Systems Research*. 25(3), 443–448.
- Agrawal, A., Gans, J. and Goldfarb, A. (2018). Prediction Machines: The Simple Economics of Artificial Intelligence. *Harvard Business Press*.
- Anderson, R. and Moore, T. (2007). The economics of information security: A survey and open questions. In *Fourth bi-annual Conference on the Economics of the Software and Internet Industries*. 19–20.
- Berente, N., Gu, B., Recker, J. and Santhanam, R. (2021). Managing Artificial Intelligence. *MIS Quarterly*. 45(3), 1433–1450.
- Berman, R. and Katona, Z. (2020). Curation algorithms and filter bubbles in social networks. *Marketing Science*. 39(2), 296–316.
- Brennen, J. S. B. and Perault, M. (2021). *Breaking Blackout Black Boxes: Roadblocks to Analyzing Platform Political Ad Bans*. Working paper.
- Breza, E., Stanford, F. C., Alsan, M., Alsan, B., Banerjee, A., Chandrasekhar, A. G., Eichmeyer, S., Glushko, T., Goldsmith-Pinkham, P., Holland, K. *et al.* (2021). Effects of a Large-scale Social Media Advertising Campaign on Holiday Travel and COVID-19 Infections: A Cluster Randomized Controlled Trial. *Nature Medicine*. 27(9), 1622–1628.
- Brynjolfsson, E., Wang, C. and Zhang, X. (2021). The Economics of IT and Digitization: Eight Questions for Research. *MIS Quarterly*. 45(1), 473–477.
- Cabral, L., Haucap, J., Parker, G., Petropoulos, G., Valletti, T. M. and Van Alstyne, M. W. (2021). The EU Digital Markets Act: A Report from a Panel of Economic Experts. *Publications Office of the European Union, Luxembourg*.
- Chiou, L. and Tucker, C. (2018). *Fake News and Advertising on Social Media: A Study of the Anti-Vaccination Movement*. Working paper. NBER.
- Choi, H., Mela, C. F., Balseiro, S. R. and Leary, A. (2020). Online Display Advertising Markets: A Literature Review and Future Directions. *Information Systems Research*. 31(2), 556–575.
- Cowgill, B. and Tucker, C. (2017). *Algorithmic Bias: A Counterfactual Perspective*. NSF Trustworthy Algorithms Working paper.
- Cowgill, B. and Tucker, C. E. (2019). Economics, Fairness and Algorithmic Bias. *Preparation for: Journal of Economic Perspectives*.
- Datta, A., Tschantz, M. C. and Datta, A. (2015). Automated Experiments on Ad Privacy Settings. *Proceedings on Privacy Enhancing Technologies*. 2015(1), 92–112.
- De Vaujany, F.-X., Fomin, V. V., Haefliger, S. and Lyytinen, K. (2018). Rules, Practices, and Information Technology: A Trifecta of Organizational Regulation. *Information Systems Research*. 29(3), 755–773.
- Edelman, B. (2009). Adverse selection in online “trust” certifications. In *Proceedings of the 11th International Conference on Electronic Commerce*. 205–212.
- Fang, X. and Hu, P. J.-H. (2018). Top Persuader Prediction for Social Networks. *MIS Quarterly*. 42(1), 63–82.
- Frank, M. W. (2008). Media Substitution in Advertising: A Spirited Case Study. *International Journal of Industrial Organization*. 26(1), 308–326.
- Fu, R., Huang, Y. and Singh, P. V. (2021). Crowds, Lending, Machine, and Bias. *Information Systems Research*. 32(1), 72–92.
- Gawer, A. and Cusumano, M. A. (2015). Platform Leaders. *MIT Sloan Management*

- Review*, 68–75.
- Goldfarb, A. and Tucker, C. (2011). Advertising Bans and the Substitutability of Online and Offline Advertising. *Journal of Marketing Research*. 48(2), 207–227.
- Hansen, K. T., Misra, K. and Pai, M. M. (2021). Frontiers: Algorithmic collusion: Supra-competitive prices via independent algorithms. *Marketing Science*. 40(1), 1–12.
- Isaac, M. (2019). *Why Everyone Is Angry at Facebook Over Its Political Ads Policy*. Retrievable at <https://www.nytimes.com/2019/11/22/technology/campaigns-pressure-facebook-political-ads.html>.
- Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S. (2018). Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*. 133(1), 237–293.
- Kopalle, P. K., Gangwar, M., Kaplan, A., Ramachandran, D., Reinartz, W. and Rindfleisch, A. (2022). Examining Artificial Intelligence (AI) Technologies in Marketing via a Global Lens: Current Trends and Future Research Opportunities. *International Journal of Research in Marketing*. 39(2), 522–540.
- Lambrecht, A. and Tucker, C. (2019). Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science*. 65(7), 2966–2981.
- Le Pochat, V., Edelson, L., Van Goethem, T., Joosen, W., McCoy, D. and Lauinger, T. (2022). An Audit of Facebook’s Political Ad Policy Enforcement. In *Proceedings of the 31st USENIX Security Symposium*. USENIX Association.
- Lefouili, Y. and Madio, L. (2022). The Economics of Platform Liability. *European Journal of Law and Economics*. 53(3), 319–351.
- Liu, Y., Yildirim, T. P. and Zhang, Z. J. (2021). Social Media, Content Moderation, and Technology. *arXiv preprint arXiv:2101.04618*.
- Ma, L. and Sun, B. (2020). Machine Learning and AI in Marketing—Connecting Computing Power to Human Insights. *International Journal of Research in Marketing*. 37(3), 481–504.
- Moore, T. (2005). Countering hidden-action attacks on networked systems. In *WEIS*.
- Nelson, J. P. (2003). Advertising Bans, Monopoly, and Alcohol Demand: Testing for Substitution Effects using State Panel Data. *Review of Industrial Organization*. 22(1), 1–25.
- Oxenford, D. (2019). *FCC Issues “Clarifications” of Political Broadcasting Public File Disclosure Requirements – Significantly More Disclosures to Be Required on Issue Ads*. Retrievable at <https://www.broadcastlawblog.com/2019/10/articles/fcc-issues-clarifications-of-political-broadcasting-public-file-disclosure-requirements-significantly-more-disclosures-to-be-required-on-issue-ads/>.
- Peer, E., Vosgerau, J. and Acquisti, A. (2014). Reputation as a Sufficient Condition for Data Quality on Amazon Mechanical Turk. *Behavior Research Methods*. 46(4), 1023–1031.
- Petratos, P. N. (2021). Misinformation, disinformation, and fake news: Cyber risks to business. *Business Horizons*. 64(6), 763–774.
- Rao, J. M. and Reiley, D. H. (2012). The Economics of Spam. *Journal of Economic Perspectives*. 26(3), 87–110.
- Sweeney, L. (2013). Discrimination in Online Ad Delivery. *Queue*. 11(3), 10:10–10:29.
- Tuchman, A. E. (2019). Advertising and Demand for Addictive Goods: The Effects of e-cigarette Advertising. *Marketing Science*. 38(6), 994–1022.

- Tucker, C. (2020). *Comment on Digital Infrastructure*, University of Chicago Press, chap. 8. 1–7. Retrievable at <http://www.nber.org/chapters/c14361>.
- Zhang, S., Mehta, N., Singh, P. V. and Srinivasan, K. (2021). Frontiers: Can an Artificial Intelligence Algorithm Mitigate Racial Economic Inequality? An Analysis in the Context of Airbnb. *Marketing Science*. 40(5), 813–820.

## A The Facebook Ad Library: SIEP Category

The Facebook Ad Library API gives access to a special category for ads related to the SIEP category. The API provides different ad characteristics and permits searches using various criteria, including keywords, ad status, targeted location and advertiser name.<sup>31</sup> Figure 11 depicts how ads are shown in the Facebook Ad Library.

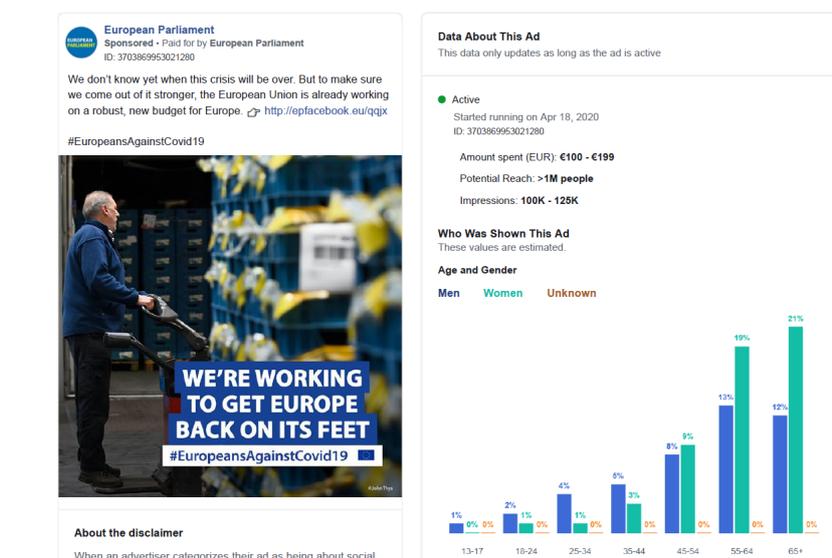


Figure 11: Platform Appearance

## B Human Classification of Random Sample of COVID-19-related Disqualified Ads

Figure 12 shows how the ad classification guidelines were presented to MTurk workers. On the left side, they can access the instructions and on the right side, we transcribe the Facebook policy guidelines and we include the ad content to review.

<sup>31</sup>The archive includes both active and inactive ads. Ads are published on Facebook, Instagram, Facebook Audience Network and Facebook Messenger. <https://www.facebook.com/ads/library/api>, April 21, 2020.

**Instructions**
✕

We're investigating Facebook's regulation of ad content. We'd love to hear how you would describe the image featured in this survey. Only one answer is possible. If you are unsure about your answer, you will be able to say so after responding to the first question. Please examine the entire ad (text and image) carefully. Thank you for your participation.

Choose the appropriate answer.

[More Instructions](#)

---

**Should this ad be categorized as special interest given the following criteria?**

- Ads made by, on behalf of, or about a candidate for public office, a political figure, a political party or advocates for the outcome of an election to public office. Or, about any election, referendum, or ballot initiative, including 'go out and vote' election campaigns.
- Ads regulated as political advertising.
- Ads about social issues in any place where the ad is being placed. Social issues are sensitive topics that are heavily debated, may influence the outcome of an election or result in/relate to existing or proposed legislation.
- Ads about social issues seek to influence public opinion through discussion, debate or advocacy for or against important topics, like health and civil and social rights.

No  
 Yes

**Submit**

Figure 12: MTurk tasks to classify COVID-19-related ads

Figure 13 shows the question that we addressed to MTurk workers after the task.

✕

Are you sure ? Whatever your answer is, you will NOT have to redo the task.

I am confident about my answer  
 I am NOT confident about my answer

Figure 13: Confirmation about MTurk Workers Choices

## C Are Workers Able to Correctly Classify Ads?

Figure 14 shows an example of one political ad that MTurk workers classify.



Figure 14: Political Ad

## D Are These Results Robust to non-English Words?

To check the robustness of our results, we searched for ads published during January to June 2020 using additional keywords which included the COVID-19-related keywords “*Handwashing*”, “*Hydroalcoholic gel*”, “*Masks*”, “*Social distancing*”, “*Stay at home*”. However, there is the possibility that our focus on the English language led us to miss ads. These keywords and translations into French, German, Italian and Spanish are presented in Table 8.

Table 8: Translation of Other COVID-19-related Keywords in Five European Languages

	Languages				
	English (1)	French (2)	German (3)	Italian (4)	Spanish (5)
Keywords	Handwashing	Lavage de mains	Händewaschen	Lavare le mani	Lavarse las manos
	Hydroalcoholic gel	Gel hydroalcoolique	Alkoholisches handgel	Gel mani alcolico	Gel antiséptico
	Masks	Masque	Masken	Maschere	Mascarillas
	Social distancing	Distanciation sociale	Soziale Distanzierung	Distanza sociale	Distanciamiento social
	Stay at home	Restez chez vous	Bleiben se zuhaus	Resto a casa	Yo me quedo en casa

Overall, the sample includes 1,165,647 different ads. The keywords “*coronavirus*” and

“*covid*” represent the large majority of ads related to the pandemic – accounting for 94.42% of the sample with the other keywords accounting for only 5.58%. In the figures, the y-axis is the proportion of disqualified ads in the samples. Figure 15 shows that COVID-19-related ads are more likely to be disqualified and especially if they are submitted by a governmental organization (Figure 16). This analysis suggest that our results are not driven by a narrow set of ads.

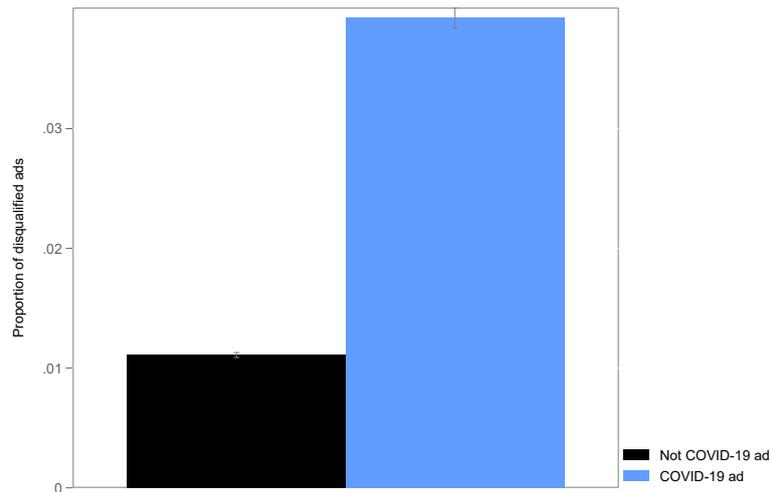


Figure 15: Proportion of Disqualified Ads by COVID-19-related ads: Sample with Large Set Keywords

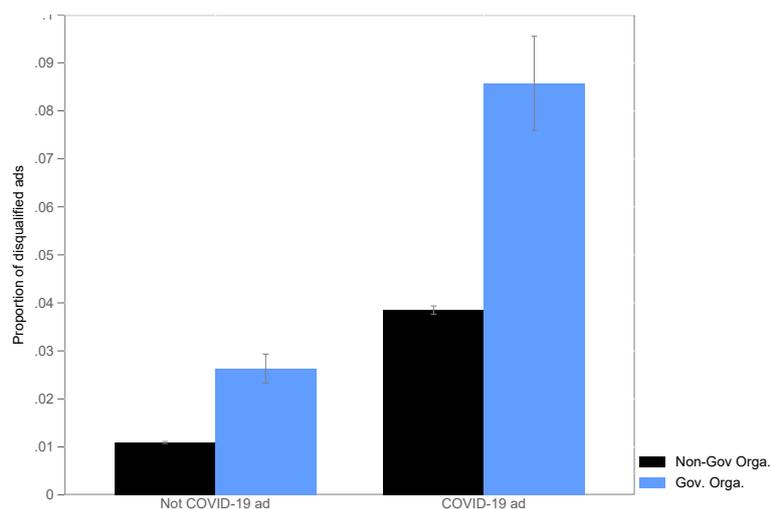


Figure 16: COVID-19-related ads Run by Governmental Organizations Are Still more Likely to be Disqualified: Sample with Large Set Keywords

## E Early Disqualification Rate

Figure 17 shows the proportion of ads disqualified for the subsample of data that excludes ads related to COVID-19 in January 2020. We split the sample between governmental and non-governmental organizations. We show that the disqualification rate was statistically similar between governmental and non-governmental organizations, suggesting that governmental organizations were not more likely to be disqualified before the pandemic and thus rule out the hypothesis that governmental organizations are more likely than other advertisers to experience a higher disqualification rate. On the contrary, it shows that they know the rules of the platform.

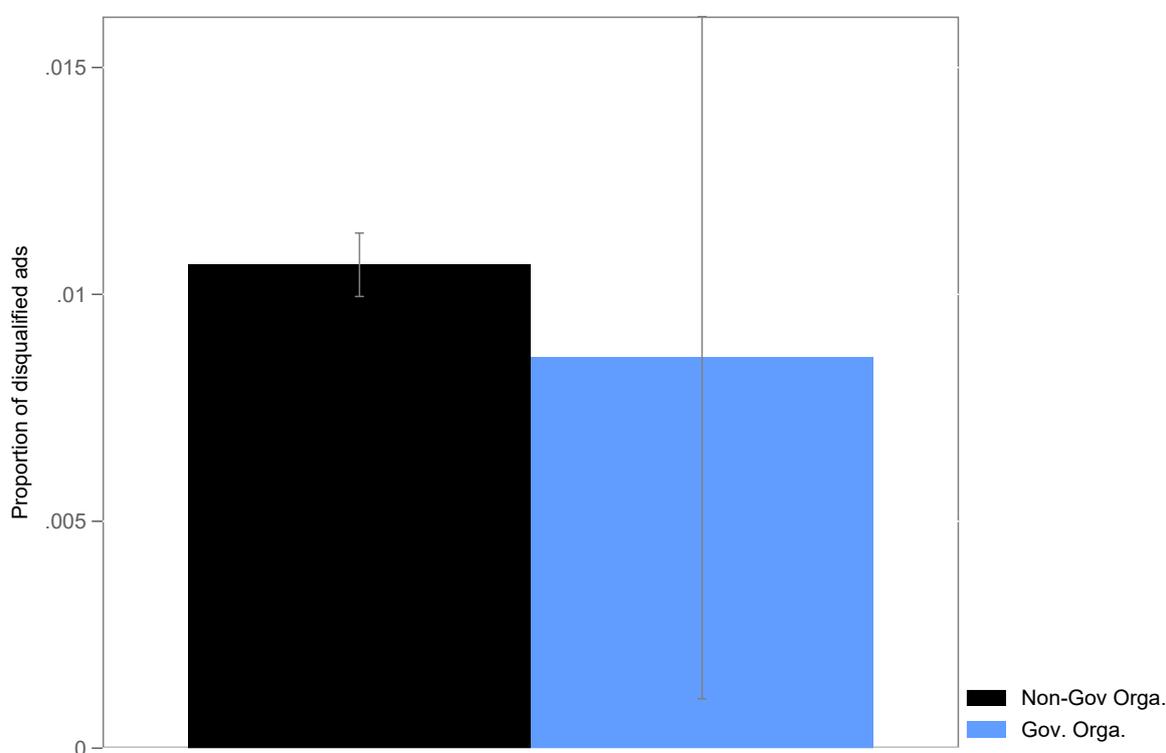


Figure 17: Do Advertisers Experience the Same Disqualification Rate Before COVID-19?

## F Robustness: Main Results with All Ads

Table 9: Estimations

	<b>Logit</b>		
	<b>Overall</b>	<b>Gov. Org</b>	<b>Non Gov. Org</b>
	(1)	(2)	(3)
COVID-19 ad	0.025 <sup>***</sup> (0.002)	0.049 <sup>***</sup> (0.004)	0.025 <sup>***</sup> (0.000)
Gov. Orga.	0.013 <sup>**</sup> (0.006)		
Europe	0.011 <sup>***</sup> (0.002)	0.016 <sup>***</sup> (0.004)	0.011 <sup>***</sup> (0.000)
Log-Likelihood			
Wald chi-squared test			
Observations	1235381	14,869	1220512

*Notes:* Significance at 1%; 5% and 10% indicated respectively by <sup>\*\*\*</sup>, <sup>\*\*</sup> and <sup>\*</sup>.