Detection and analysis of online issue and political ads

Vera Sosnovik

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

École doctorale : MSTII - Mathématiques, Sciences et technologies de l'information, Informatique
Spécialité : Mathématiques et Informatique
Unité de recherche : Laboratoire d'Informatique de Grenoble

Détection et analyse des publicités qui adresse des enjeux sociaux et des publicités politique en ligne

Detection and analysis of online issue and political ads

Présentée par :

Vera SOSNOVIK

Direction de thèse :

Patrick LOISEAU CHARGE DE RECHERCHE INRIA, INRIA Saclay	Directeur de thèse
Oana GOGA CHARGEE DE RECHERCHE HDR, CNRS DELEGATION ALPES	Co-directrice de thèse

Rapporteurs :

WALTER RUDAMETKIN
PROFESSEUR DES UNIVERSITES, UNIVERSITE DE RENNES
KEVIN HUGUENIN
FULL PROFESSOR, UNIVERSITE DE LAUSANNE

Thèse soutenue publiquement le **4 septembre 2023**, devant le jury composé de :

WALTER RUDAMETKIN PROFESSEUR DES UNIVERSITES, UNIVERSITE DE RENNES	Rapporteur
KEVIN HUGUENIN FULL PROFESSOR, UNIVERSITE DE LAUSANNE	Rapporteur
JUHI KULSHRESTHA ASSISTANT PROFESSOR, AALTO YLIOPISTO	Examinatrice
GILLES BASTIN PROFESSEUR DES UNIVERSITES, UNIVERSITE GRENOBLE ALPES	Président
PAOLO FRASCA CHARGE DE RECHERCHE, CNRS DELEGATION ALPES	Examinateur

Invités :

PATRICK LOISEAU
CHARGE DE RECHERCHE HDR, INRIA CENTRE GRENOBLE-RHONE-ALPES
OANA GOGA
CHARGEE DE RECHERCHE HDR, CNRS DELEGATION ALPES

Abstract

Online political advertising has become the cornerstone of political campaigns. The budget spent solely on political advertising in the U.S. has increased by more than 100% from $ 700 million during the 2017-2018 U.S. election cycle to $ 1.6 billion during the 2020 U.S. presidential elections. Naturally, the capacity offered by online platforms to micro-target ads with political content has been worrying lawmakers, journalists, and online platforms, especially after the 2016 U.S. presidential election, where Cambridge Analytica has targeted voters with political ads congruent with their personality.

To curb such risks, both online platforms and regulators (through the DSA act proposed by the European Commission) have agreed that researchers, journalists, and civil society need to be able to scrutinize the political ads running on large online platforms. Consequently, online platforms such as Meta and Google have implemented Ad Libraries that contain information about all political ads running on their platforms.

The thesis consists of three contributions related to the online political advertising problems. The first project investigates whether we can reliably distinguish political ads from non-political ads. We take an empirical approach to analyze what kind of ads are deemed political by ordinary people and what kind of ads lead to disagreement. Our results show a significant disagreement between what ad platforms, ordinary people, and advertisers consider political and suggest that this disagreement mainly comes from diverging opinions on which ads address social issues. Overall our results imply that it is important to consider social issue ads as political, but they also complicate political advertising regulations.

In the second project, we focus on political ads that are related to policy. Understanding which policies politicians or organizations promote and to whom is essential in determining dishonest representations. We propose automated methods based on pre-trained models to classify ads in 14 main policy groups identified by the Comparative Agenda Project (CAP). We discuss several inherent challenges that arise.

Finally, we analyze policy-related ads featured on Meta platforms during the 2022 French presidential elections period.

In the final contribution we propose a set of practical benchmarks to evaluate the "goodness" of political ad definitions. The benchmarks aim to assess whether the definitions can capture a set of truly problematic ads (the true positives), such as ads with divisive messages across demographic groups, and the ability to not capture a set of ads that only have humanitarian scopes (the false positives). We evaluate two definitions from online platforms and two definitions from policymakers based on our benchmarks. Our results show that definitions that only cover ads from/about political actors, and elections miss the highest percentage of advertisements that are divisive across different demographic groups.

Résumé

La publicité politique en ligne est devenue la pierre angulaire des campagnes politiques. Le budget consacré uniquement à la publicité politique aux états-Unis a augmenté de plus de 100 %, passant de 700 millions de dollars lors du cycle électoral américain de 2017-2018 à 1,6 milliard de dollars lors des élections présidentielles américaines de 2020. Naturellement, la capacité offerte par les plateformes en ligne de micro-cibler les publicités à contenu politique inquiète les législateurs, les journalistes et les plateformes en ligne, en particulier après l'élection présidentielle américaine de 2016, où Cambridge Analytica a ciblé les électeurs avec des publicités politiques conformes à leur personnalité.

Pour limiter ces risques, les plateformes en ligne et les régulateurs (par le biais de la loi DSA proposée par la Commission européenne) ont convenu que les chercheurs, les journalistes et la société civile doivent être en mesure d'examiner les publicités politiques diffusées sur les grandes plateformes en ligne. Par conséquent, les plateformes en ligne telles que Meta et Google ont mis en place des bibliothèques d'annonces qui contiennent des informations sur toutes les publicités politiques diffusées sur leurs plateformes.

La thèse se compose de trois contributions liées aux problèmes de la publicité politique en ligne. Le premier projet étudie si nous pouvons distinguer de manière fiable les publicités politiques des publicités non politiques. Nous adoptons une approche empirique pour analyser quels types de publicités sont considérées comme politiques par les gens ordinaires et quels types de publicités conduisent à des désaccords. Nos résultats montrent un désaccord significatif entre ce que les plateformes publicitaires, les gens ordinaires et les annonceurs considèrent comme politique, et suggèrent que ce désaccord provient principalement d'opinions divergentes sur les publicités qui traitent des problèmes sociaux. Dans l'ensemble, nos résultats impliquent qu'il est important de considérer les publicités à caractère social comme politiques, mais ils compliquent également la réglementation de la publicité politique.

Dans le deuxième projet, nous nous concentrons sur les publicités politiques liées à la politique. Comprendre quelles politiques les politiciens ou les organisations promeuvent et auprès de qui est essentiel pour déterminer les représentations malhonnêtes. Nous proposons des méthodes automatisées basées sur des modèles pré-entraînés pour classer les publicités dans 14 principaux groupes de politiques identifiés par le Comparative Agenda Project (CAP). Nous discutons de plusieurs défis inhérents qui se présentent. Enfin, nous analysons les publicités liées aux politiques présentées sur les plateformes Meta pendant la période des élections présidentielles françaises de 2022.

Dans la contribution finale, nous proposons un ensemble de repères pratiques pour évaluer la "qualité" des définitions de la publicité politique. Les benchmarks visent à évaluer si les définitions peuvent capturer un ensemble d'annonces vraiment problématiques.

Contents

Chapter 1

Introduction

1.1 Misuse of online political ads

Traditionally political parties have used manifestos to communicate the set of policies they announce they would implement if elected [93] and promoted their political agendas through mass media. With the emergence of online advertising platforms, online ads have become one of the main communication channels for political campaigners. During the 2020 US election cycle, 18% of political marketing spending went to online advertising, compared to 3% during the 2016 election cycle [36]. Moreover, online advertising spending by parties increased from 24% to 43% of advertising budgets between the UK general elections of 2015 and 2017 [91].

In several countries, an "industry of political influence" is setting up ways to profile electorates and disseminate "micro-targeted" messages through various techniques for narrow groups of voters according to their profiles. Thus, the integrity of elections has been increasingly threatened in recent years by these covert practices.

They mobilize social media such as Facebook, which, thanks to the data they hold on their users, has acquired a strong targeting capacity allowing them to address messages to the targeted audience and to refine the content of these messages according to their interests.

Besides the low cost, the key appeal of online micro-targeted advertising for political campaigners comes from the fact that they can communicate a more diverse set of information (than traditional mass media), and they can target subgroups of voters with information that is relevant to them. However, many researchers and civil societies are firing alarms that targeting technologies are also allowing the emergence of an "industry of political influence" [7] where political advertisers can select very narrow groups of vulnerable people and tweak their messages to maximize their influence [77].

The significant growth of digital political advertising and the lack of regulation and supervision led to a misuse of the technology. One of the most known examples is the Cambridge Analytica data scandal [61]. They used the personal data of more than 50 million Facebook users during Donald Trump's 2016 US presidential election campaign and targeted citizens with ads tailored to their personalities. Another example involves Canadian data firm AggregateIQ, which worked with two pro-BREXIT campaigns (VoteLeave and BeLeave) during the 2016 United Kingdom European Union membership referendum [53]. AggregateIQ is known for targeting specific groups of people with false or half-truth statement ads on Facebook.

1.2 Governance of online political advertising

1.2.1 Platforms' restrictions

Ad platforms have put forward several measures to mitigate risks and allow for public scrutiny of ads. Twitter and TikTok decided to ban political ads altogether. Google and Facebook allow political ads, but advertisers are subject to a higher degree of scrutiny and limitations. On Google, advertisers can only use geographic location, age, gender, and contextual targeting to target political ads. Facebook does not restrict the micro-targeting of political ads. Advertisers, however, need to verify their account (by showing proof of identity or a public listing of their business [32]) and are only allowed to send political ads to users that reside in the same country as them. Moreover, advertisers have to self-declare when their ads are political, and all political ads sent on the platform appear in the Facebook Ad Library where the civil society can further scrutinize them [30].

1.2.2 Government's regulation

Governments on their side have been working on projects to regulate and monitor online (political) advertisements. One of them is the Digital Services Act (DSA) [27].

DSA is a set of rules to apply across the European Union. The main goal of it is to create a safer digital space in which fundamental rights are protected and to establish a level playing field to foster innovation, growth, and competitiveness. DSA designates rules for social networks, content-sharing platforms, online marketplaces, etc. the Digital Services Act contains a part about laws related to online political advertising. Such as online platforms and search engines with more than 45 million

monthly EU users are obliged to have ad libraries that contain information about *all ads* running on their platforms and information about how the ads were targeted.

European Democracy Action Plan (EDAP) strengthens media freedom and promotes fair elections. The action plan consists of several packages of rules aiming to build sustainable democracies across the European Union. One of the parts of the EDAP contains measures for regulating online and offline political advertising. They insist that all digital sponsored political content should have information about the identity of the sponsor of the political advertisement and the entity ultimately controlling the sponsor available in a clear way.

1.3 Contributions

In this thesis, we focus on analyzing an online political advertisement. We also explore different algorithms for the detection and classification of digital political ads. Moreover, we propose benchmarks for the evaluation of definitions of political advertisement.

1.3.1 Detection of political ads

Measures from both ad platforms and governments are positive developments. However, all of them implicitly rely on the assumption that *one can reliably distinguish political ads from non-political ads.*

We take an empirical approach to test this assumption by analyzing the characteristics of ads deemed political by ordinary people, the characteristics of ads that lead to disagreement, and whether there are differences between what advertisers consider political and what ordinary people consider political.

Our analysis is based on a dataset from ProPublica [75] that contains 55k Facebook ads received by U.S. residents, labeled by at least one volunteer as political, and that received three or more votes. The dataset was collected by a browser extension that collects the ads users see when they browse their Facebook timeline and allows users to label whether the ads they see are political.

First, we investigate whether ad platforms, volunteers, and advertisers agree on which ads should be considered political. All ad platforms agree that ads from or about political actors and ads about elections and voting should be considered political. However, only Facebook and TikTok consider ads about *social issues* (such as climate change or immigration) as political. Our results show that volunteers disagree on whether an ad is political for more than 50% of the ads in the dataset, and

only 83% of the ads labeled as political by advertisers are also labeled as political by a majority of volunteers. Hence, the fundamental assumption that we can clearly distinguish political from non-political ads does not hold, since there is no consensus even on what constitutes a political ad, and volunteers and advertisers label different sets of ads as political.

Next, we analyze the characteristics of ads that are labeled as political by volunteers and advertisers in the ProPublica dataset, which can be useful to inform the debate on definitions of political ads. To that end, we gathered data about the advertisers sending political ads and the content of their ads. We hired Prolific users to annotate 2300 ads with the political or social issues the ad is referring to. Our analysis shows that a wide range of advertisers (from political actors to NGOs and businesses) are posting political ads on Facebook and that ads about social issues account for a large fraction of the ads labeled as political; hence emphasizing the importance of including such ads in political ads definitions. Our analysis also shows that the ads labeled as political by volunteers and advertisers are very diverse. We see ads with a clear political message from advocacy groups (e.g., ads addressing abortion issues in the U.S.); but also ads from NGOs that address humanitarian issues and do not seem to directly or indirectly impact U.S. elections or legislation (e.g., ads asking for donations for ending world hunger). As political ads may be subject to higher restrictions, this questions whether it is desirable that the same restrictions apply to both types of ads. More generally, this emphasizes the need to account for the diversity of political ads in devising regulations.

We finally analyze the ads that lead to disagreement among volunteers and between volunteers and advertisers. We first observe that advertisers mislabel some ads as either political or non-political (according to the Facebook ToS).

Then we find that advertisers seem to underreport ads (that are considered political by volunteers) about social issues, especially the economy and civil and social rights. Volunteers seem to underreport ads (considered political by advertisers) from advertisers such as NGOs and charities, and about social issues, especially civil and social rights and health. Part of the problem may be that the definition of ads about social issues may be too broad and vague, which leads to being interpreted in different ways by people. This also raises the question of whether *all* ads related to social issues should be considered political, and if not, how to filter social issue ads that are not political.

Because of the high volume of ads, enforcement mechanisms need to rely on automated machine learning (ML) algorithms to detect political ads. However, it is not

clear how one should train and evaluate such models since there is disagreement on which ads are political (i.e., the positive examples). To investigate that, we train four classifiers with different groups of positive examples (coming from advertisers and volunteers). We test how they perform over various groups of political ads with varying degrees of disagreement. While all classifiers achieve high accuracy in detecting ads everyone agrees are political; their accuracy drops on ads that only a few find political.

Another important question is whether (and to which extent) models trained with labels from advertisers would declare as political the same ads as models trained with labels from volunteers (i.e., reliable detection of political ads). Theoretically, if ads labeled as political by advertisers and volunteers are representative of political ads in general, the resulting models should declare the same ads as political. Our results show that the overlap between different models is relatively high (ranging from 82% to 97%), but that discrepancies in the input data transfer to discrepancies in the output data. This suggests that existing labeled datasets are not providing a representative set of political ads needed to build reliable detection schemes.

Overall, our work suggests that, given the complexity of deciding which ads are political, it would be beneficial to have ad libraries that contain *all* ads running on the platform, not only ads deemed political by the ad platform. Following this work, we issued a statement together with civil societies asking for "*Universal advertising transparency by default*" that we submitted to the DSA consultation [35]. However, this crucial first step is not enough because political ads are also subject to higher restrictions; hence, we still need to detect political ads reliably. We hope this study can help policymakers to define political speech and decide on appropriate restrictions and ad platforms to set infrastructures for detecting political ads. The results of this work were published in the proceedings of International World Wide Web Conference 2021, and were presented at the respective conference.

1.3.2 Classification of policy-related political ads

We focus on methods for *detecting policy-related political ads*. There are a number of reasons why identifying policy-related political ads is important: (i) *political communication*– makes it possible to identify how political candidates and parties represent themselves and on which policies they focus their attention; (ii) *mandate accountability*–check, once elected, whether elected officials respected the policy pledges they advertised during elections (accountability is central to democratic theory [92]); (iii) *influence on deliberation*–mandate theories assume that voters are rational and they

decide for whom to vote based on a careful consideration of available information [55]. In practice, the deliberation process is more complex and is often based on emotions, convictions, and experiences [88]. Policy-related ads are interesting in both "rational voter" and "emotional voter" models. Micro-targeting of policy-related ads could lead to some users being overly exposed to ads about specific policy issues (e.g., immigration), which might trigger strong emotions. In contrast, other voters might not get sufficiently exposed to any policy-related ads, which could lead to information incompleteness.

We use the CAP codebook as the underlying theoretical basis seems more suitable in the context of political micro-targeted ads [12].

For the analysis, we gathered more than 96k political ads from the Meta's Ad Library that appeared between 1 Jan and 14 June 2022. To gather labeled data, two experts annotated 431 ads with the relevant CAP categories. To complement this dataset, we used Prolific [73] and Qualtrics [76] to post assignments for annotating ads, and we gathered labels for 4 465 ads. We observe only fair agreement (kappa>0.3) between Prolific users and experts. We show disagreement mainly happens on ads that are related to more than two policy categories, hence, disagreement is linked to the text complexity of real-world ads.

We implemented several machine learning (ML) models to classify ads in the relevant CAP categories based on both traditional supervised models and pre-trained language models based on BERT that exploit as training data from CAP and annotations from Prolific users. Our best configuration is able to achieve a micro average F1 score of 0.60 over a balance test set. The accuracy varies drastically depending on the policy category and ranges from a 0.19 F1 score for "Social policy" to a 0.78 F1 score for "Environment". The differences are explained by the disagreement present in the training data and the labeling complexity of real-world ads.

Finally, to show the practical usefulness of the classifier we developed, we analyze how policy attention varied across candidates and different demographic groups during the 2022 French Presidential election. Overall, we see big variations in policy attention across demographic groups, with women over-targeted with ads about "Health", young users (ages 13-24) over-targeted with ads about "Law and crime" and users aged over 55 over-targeted with ads about "Immigration". This kind of imbalance could reinforce gender and age stereotypes, and may deprive users from relevant information that might be important in their voting deliberation.

The results of this work were published [85] in the proceedings of International World Wide Web Conference 2023, and were presented at the respective conference.

1.3.3 Benchmarks

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. 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:

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.

We assess the quality of the four definitions of political advertisement from different sources (from social media platforms and official government documents).

To assess the quality of the definitions, we set up three studies on Prolific. In the first one, we ask workers to label the political ads from Meta Ad Library according to one of the definitions. In two others, we ask questions about the content of an ad.

We find that the definition that only covers ads from/about political actors and elections misses the highest percentage of divisive advertisements across different demographic groups. Moreover, we also indicate that definitions that include ads about social issues cause the most significant disagreement among workers.

While there might be other dimensions across which we might want to benchmark definitions of political ads that are not covered here, we hope our paper is a start for setting scientific approaches to assess definitions of political ads

1.4 Organization of the thesis

The thesis is organized as follows. Chapter 2 presents the state of the art of studies of online targeted advertising, online political advertising, online political content analysis and auditing Meta Ad Library. Chapter 3 describes advertising process on

Meta and its difference with political advertising process. Chapter 4 shows our work on detection political ads. Chapter 5 presents our work on classification of political ads. Chapter 6 describes benchmarks for evaluation political ads definitions. We conclude in chapter 7.

Chapter 2

State of the Art

2.1 Introduction

The growth of online political advertising provokes a series of research focusing on analyzing it in computer and political science communities. Firstly, this chapter provides an overview of studies on online targeted advertising focusing on the discrimination problems. Next, we present state of the art of studies on online political advertising and online political content analysis. Finally, we review Meta Ad Library auditing studies.

2.2 Studies of online targeted advertising

Online targeted advertising in recent years became the primary source of attracting customers and outperformed offline ads campaigns [39]. Online ads are more affordable because they do not calculate cost-per-thousand-impressions (CPM) but cost-per-click (CPC). Moreover, online platforms allow advertisers to target users with specific features [16].

However, targeting can also be a tool for discrimination against various groups of people. Non-governmental organization ProPublica showed that Facebook allows advertisers to exclude people by age, race, and gender in job and housing ads [4, 48, 94]. In response to these studies, Facebook restricted targeting options in housing, employment, and credit advertisements [63].

Nevertheless, this measures are not enough to be sufficient. Speicher et al. [86] show that they are sill opportunities to discriminate sensitive groups of people. They describe how malicious advertisers can use PII-based, look-alike audience and attribute-based targeting for it. For instance, PII data is available from many public sources such as voter records, criminal history records and data brokers. Some of

these sources can have sensitive attribute information. Thus, for PII-based targeting, advertisers can upload one of the record with filtering based on the race/age/gender of the person. What is also important is that this process is not transparent for a platform. They also show that expanding the audience based on the biased list, so Facebook will propagate this bias. Finally, advertisers can also have a possibility of discrimination with attribute-based targeting. Advertisers still can use free-from attributes, that can provide sensitive information.

Series of studies analysed discrimination in ad delivery algorithms [1, 46, 49]. Ali and Sapiezynski et al. [1] demonstrate the significant skew in delivery along gender and racial lines for advertisement about job and housing opportunities despite neutral targeting parameters. Through their experiments, they showed that images, rather than the texts, titles or authors, influence the delivery of an advertisement to a certain group more than anything else.

Imana et al. audited Facebook and LinkedIn algorithms for delivering ads about different types of jobs [46]. They set up three series of experiments with a low-skilled job (delivery driver), a high-skilled job (software engineers), and a low-skilled but popular job among the audience (sales associates) on these two platforms. In all experiments, Facebook's ad delivery algorithm is skewed by gender, even when advertisers intended to target gender balanced audience. Conversely, the LinkedIn algorithm did not show gender imbalance while displaying ads. Their findings show that the Facebook algorithm may violate anti-discrimination laws [26]. Another study was focus on gender discrimination only in STEM career ads [49]. They sent ads on Facebook about job opportunities in STEM and did not use gender as a targeting parameter. Their results show that fewer women than men saw these ads. The authors find out that the main reason for this discrimination is that displaying an advertisement for younger women is more expensive than for others.

We overview these works to show the problems related to online targeting advertising and emphasize that it requires more regulation and audition. However, in our work, we focus only on online political advertising.

2.3 Studies of online political advertising

Cases of missuses of online political advertising aroused increased interest in this topic not only among social and political scientists but also among the computer science community.

Many studies analyse sponsored political content from the Meta ad library []. A few early works have used *manual labeling* to encode political ads according to various characteristics and analyze the results. Calvo et al. [14] collected 14 684 ads from six parties during Spain's 2019 general election. They manually coded 1 743 ads according to 9 topics of interest to understand how much money different parties spend on promoting different topics. Party promotion was the topic on which all six parties spent most of the budget. Dobber et al. [24] analyzed the electoral promises Dutch political parties were making during the 2019 European elections. The authors collected and labeled 362 ads according to the CAP codebook. Their analysis showed that political campaigns promoted electoral promises only to small groups of people and concluded that this is problematic from a democratic accountability perspective. These sorts of questions are what motivated us to propose automated methods that enable robust and large-scale analyses. Study of Flower et al. [38] analyzed both offline and online political advertising. It showed, by comparing Facebook posts from 7 056 candidates and T.V. ads from 1 274 candidates in the 2018 U.S. mid-term election, that Facebook posts are used for a more diverse range of goals–such as fundraising than are TV ads.

Capozzi et al. [15] focused on populist parties' political ads during the 2019 European Parliamentary election. They analyzed the differences and similarities in their content and the reached audience between different countries. They also showed that even if populist parties represent only 20% of the total spending on sponsored political content, they score 40% of total impressions.

Finally, Gitomer et al. [40] analyzed location targeting of political ads in Germany. In our work, we dig deeper and analyze if different policy categories receive more/less attention across different regions in France.

Manual labeling of large amounts of ads is time-consuming and costly, a few recent works have proposed methods to *automatically label ads*. Baviera et al. [9] used the Key-phrase Digger algorithm [65] to detect the main terms in the texts of the 14 684 Facebook ads. They found that the main aim of the ads was to mobilize voters. This work is orthogonal to ours as it provides a less comprehensive but more focused perspective on topics discussed. Coelho et al. [18] analyze the differences between political ads on Facebook in English and Spanish during the 2020 U.S. presidential election. They used a two-layer fully connected neural network to predict topics for the ads and showed that Spanish speakers received more natural and informative ads.

Regarding video content, Baskota et al. [8] proposed methods to classify the tone in political videos and They classified videos by extracting the text from audio and

key-frames and using it as the features for classifiers. The results showed that SVM with handcrafted features and oversampling performed best in the test set. Banerjee et al. [6] proposed methods to differentiate political campaign ads from other online video ads by using both textual and non-textual features.

2.4 Studies of online political content analysis

More broadly, a lot of research in both political and communication sciences has analyzed online political content on social media. Because of space constraints we only discuss works related to policy analysis.

Rusell et al. [78] examined what women in the U.S. Congress discuss on Twitter. She collected 113 112 tweets from verified senator's accounts and trained students to *manually label* them according to 20 major topics from U.S. Policy Agenda Project coding scheme. The results showed that congresswomen post on Twitter about diverse topics and do not focus only on women-related issues.

The biggest problem when building *automated methods to label political text* is the lack of labeled data. Terechshenko et al. [89] propose to use transfer learning and showed that RoBERTa achieved the highest accuracy score of 61% when trained on the CAP bills dataset and tested on the CAP New York Times headlines. In this paper we showed that transfer learning from CAP bills to ads results in very low accuracy. Hemphill et al. [44] investigated policy attention among different U.S. congress members on Twitter. They manually labeled 59k tweets according to the CAP scheme. Using logistic regression with bag of words they achieved a 0.79 F1. They found that the proportion of congress members' tweets about policy issues stayed stable. The paper does not provide any details on the annotation process and does not show the accuracy across different policy categories. In fact, some of the results could be invalidated if the recall differs for different policy categories as we show it happens in the context of political ads. Nevertheless, our work provides a deeper understanding on the limitations of using state-of-the-art automated methods to label policy-related ads that could potentially apply to other social media texts. Finally, Jackson et al. [47] proposed to use a lexicon-based approach to built a list of language cues for nine political topics to deal with the lack of training data. The authors used transcripts of primary debates to obtain seed words for the lexicon and used 29k Facebook posts (from Republican, Democratic, third parties presidential candidates) and 98M tweets (from presidential candidates and people who mention them) to revise and test the lexicon. The authors evaluated the method over 500 labeled texts and they achieved

an accuracy of over 85% for eight out of nine categories. Gupta et al. [42] used a supervised approach for classifying ads into different categories such as advocacy, attack, image, and issue; but without investigating the precise issue discussed. The authors manually labeled 5 231 Tweets and 4 434 Facebook posts which they used to build a BERT classifier that achieves an accuracy of 83%. Overall, there have been several related works on analyzing political content, however, none of them provides the solid foundations we provide for analyzing policy-related ads that goes from having the right codebooks, investigating difficulties in annotation and understanding which language models configurations are most suitable for supporting such nuanced classification. In our work we also the first to analyze policy attention in political ads at large-scale and show imbalances across demographic groups.

2.5 Meta Ad Library auditing

After Meta opened Ad Library, researchers, journalists, and civil societies accessed the sponsored political content and its insides. However, this tool needs to be more transparent and lead to further investigation. Marcio et al. [82] wanted to discover how many political ads were missed from the Brazil Ad Library. To monitor and analyze political messages, the authors created a browser extension to collect ads from Facebook timelines. They implemented several supervised classifiers to detect political ads during the election period. The results have shown that Meta Ad Library has an equivalent number of declared and undeclared political ads. Pochat et al. [72] focused on the work of Meta's political ads reinforcements. The authors showed that even if Meta algorithms are doing relatively well and can detect more than 40% unlabeled ads in less than 24 hours, users are still exposed to infringing content since these unlabeled political ads got more than two billion impressions. Edelson et al. [25] conduct a security audit for the U.S. Ad Library. They discovered that 54.6% of pages with political ads included in the Ad Library never provide a disclosure string. Despite the promise to keep any political ad in the library for seven years, some ads were deleted from the archive. Overall, Meta Ad Library is an excellent tool that helps reduce the misuse of political ads, it has design and implementation flaws.

2.6 Definitions of what is political

Several works studied what people think is political [34, 43]. Fitzgerald [34] set up several experiments to identify what topics people consider political and showed

that, similar to our study, it is complicated to defined what topics are political. The experiment included 33 different topics such as education, poverty, national parks and space exploration. The participants achieved a 95% agreement that diet pills is not a political topic. However, the mean percentage of respondents who view a topic as political is 42%. Hansford et. al. [43] designed an implicit association test featuring the Supreme Court and Congress. The results showed that people perceived the Supreme Court as less implicitly political than Congress. On a more theoretical side, Sartori [79] considered the question of the autonomy of politics. The author concluded that the current situation of politics is reflected in three different ways: outright extinction, autonomy or weakening, which leads to different ways of perceiving, identifying, and defining politics. Warren [96] proposed that the concept of politics should help to clarify normative interests in politics, that the definition of politics should embrace everyday understandings of politics, and serve explanation. He suggested that politics can be define by two attributes: power and conflict.

Chapter 3

Background

3.1 Political Advertising on Meta

To become an advertiser on Meta, the requirements are simply to have a Facebook account and provide a valid payment method. However, the process differs for advertisers intending to promote ads related to social issues, elections, or political figures. Anyone who wants to do it should complete the following steps.

Advertisers must complete authorization by providing their photo id, two official documents, or a notarised form of the country where ads will be run. They also must have two-factor authentication enabled, and advertisers should be an admin of the page that send ads. They are allowed to send sponsored political content only in the countries they reside. However, they are an exception for organizations with a membership of three or more sovereign states bound together by a treaty that can send ads about only specific social issues across the European Union, the United Kingdom, and the United States. The content of the ads must not include electoral, political, or legislative topics.

Finally, on all ads about social issues, elections, or political actors, advertisers must put a "Paid for by" disclaimer with the information about who sponsored this advertisement. This "Paid for by" disclaimer appears on the top of the ad frame, after the advertiser's name. Finally, Facebook ads the political ads to their Ad Library [30].

3.2 Political ads' definitions

This section presents the political ads definitions proposed by various platforms and lawmakers.

Facebook defines political ads as [32]:

> *"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" or election campaigns; or*
> *About social issues in any place where the ad is being placed; or*
> *Regulated as political advertising."*

The social issues list depends on where the ad was published. In the European Union, the following social issues should be regulated as political ads: Civil and social rights, crime, economy, environmental politics, health, immigration, political values and governance, and security and foreign policy. Broadly speaking, the Facebook definition covers three categories of ads: ads from/about political actors, ads about elections, and ads about social issues.

Twitter defines political ads as [95]:

> *Ads with political content: that references a candidate, political party, elected or appointed government official, election, referendum, ballot measure, legislation, regulation, directive, or judicial outcome; ads that contain references to political content, including appeals for votes, solicitations of financial support, and advocacy for or against any of the above-listed types of political content, are prohibited under this policy; as well as ads of any type by candidates, political parties, or elected or appointed government officials.*

Broadly speaking, the definition only covers ads from/about political actors and ads about elections and does not cover ads about social issues.

Google put political ads and election ads into two separate categories. There is no restriction for political ads from Google, except it should respect the country's laws. The definition of election ads is different according to region [41]. EU Election Ads are ads that feature any of the following:

> *A political party, current elected officeholder, or candidate for the EU Parliament;*
> *A political party, current officeholder or candidate for an elected national office within an EU member state. Examples include members of a national parliament and presidents that are directly elected; or*
> *A referendum question up for vote, a referendum campaign group or a call to vote related to a national referendum or a state or provincial referendum on sovereignty.*

Hence, Google's definition does not cover social issue ads.

TikTok defines political ads as [17]:

> *Ads that promote or oppose a candidate, current leader, political party or group, or issue at the federal, state, or local level — including election-related ads, advocacy ads, or issue ads* [17].

Hence, TikTok's definition covers social issue ads besides ads from/about political actors and ads about elections.

European Commission The European Commission is working on a set of laws and measures that are part of the European Democracy Action Plan (EDAP) [19] that aim to counter disinformation and promote free and fair elections. As part of EDAP, in 2020, the European Commission proposed a text to regulate political advertising "Proposals on the Transparency and Targeting of Political Advertising" [20]. The document presents the following definition:

> *Political advertising means the preparation, placement, promotion, publication or dissemination, by any means, of a message:*
> *(1) by, for or on behalf of a political actor, unless it is of a purely private or a purely commercial nature; or*
> *(2) which is liable to influence the outcome of an election or referendum, a legislative or regulatory process or voting behaviour.*

The European Commissions' definition distinguishes from the Facebook and Twitter definition by insisting on the outcome of the ads: "liable to influence their outcome" rather than what the ads are talking about (e.g., social issues).

Later on, in 2022 **European Parliament's** Committee on Culture and Education proposed an amendment to the European Commission's definition [69]. They formulated the new definition as:

> *Political Advertising means the preparation, placement, promotion, publication or dissemination, provided for remuneration, which may include a benefit in kind, by any means, of a message:*
>
> 1. *by, for or on behalf of a political actor; or*
> 2. *related to an election or referendum, a legislative or regulatory process or voting behaviour at European, national, regional, local or at a political party level, and designed to influence their outcome.*
>
> *It shall not include message of a purely private or a purely commercial nature, purely journalistic content, or political views expressed in the programs of audiovisual linear broadcasts or published in printed media without direct payment or equivalent remuneration.*

Finally, the **U.S. Federal Election Commission** does not use the term political adverting. However, it makes a difference between advertising about candidates or

elections and advertising related to public policy issues without mentioning candidates [81]. Issue ads are not regulated.

Overall, while there are similarities, we can see that there is no agreement over what should be a political ad neither between platforms nor across lawmakers. Differences in definitions impact what ads get labeled as political, and consequently, what ads would be restricted. For example, while Twitter banned political advertising in 2019, its definition only covered ads from or about politicians and ads about elections, and it did not cover social issue ads. Given the importance of regulating political advertising, it is essential to work towards a good definition for political ads.

Chapter 4

Detection of political ads

This chapter covers the work in the paper: WWW'21 [83].
This work was done in collaboration with my supervisor Oana Goga (CNRS).

In this chapter, we investigate whether there are significant differences between ads labeled as political by advertisers and ads labeled as political by a group of volunteers. In section 4.1, we describe the data we used. Section 4.2 examines whether there is consensus among ad platforms, volunteers, and advertisers on what ads should be categorized as political. We conduct an analysis of the attributes of ads that are identified as political by volunteers and advertisers in section 4.3. This analysis aims to contribute to the ongoing discourse surrounding the definition of online political ads. We examine the ads that leads to disagreement among volunteers, as well as between volunteers and advertisers in section 4.4. In 4.5 we test how classifiers perform over various groups of political ads with varying degrees of disagreement. Section 4.6 describes the server we created to explore political ads. Lastly, we analyze how task design influences people's decisions on what ads are political in the section 4.7.

4.1 Datasets

For our analysis we use the following two datasets of ads that users have received on their Facebook timeline:

ProPublica dataset ProPublica, an investigative journalism organization, has developed a browser extension that collects the ads users are receiving on Facebook and allows users to label whether the ads they are seeing are political or not [75]. The extension is currently maintained by the NYU Online Political Transparency Project [60]. While ProPublica was not able to make available all the ads it has collected, it shared with us **all the ads for which at least one user has labeled**

Table 4.1: Number of ads in the ProPublica and the AdAnalyst datasets, and percentage of ads with official "Paid for by" political disclaimer.

	All ads	Official political	Official non-political
ProPublica	54.6k	50.8k (93%)	3.8k (7%)
AdAnalyst	9k	2%	98%

it as being political, as well as **all the ads that have the "Paid for by" disclaimer** (i.e., the official political ads that have been declared as such by advertisers). This dataset is valuable because it provides us with a unique view of which ads are considered political by "ordinary" people/volunteers. To our knowledge, there are no studies of such data.

For this study, we only kept ads with at least three votes (either political or non-political) and that were received between June 2018 and May 2020; resulting in a dataset of 54.6k ads coming from 7530 advertisers. The median number of votes per ad after filtering is 5. We call the ads that have the "Paid for by" disclaimer the **official political ads** and the ads that do not have the disclaimer the **official non-political ads**. Table 4.1 shows the number of ads in the ProPublica dataset as well as the fraction of **official political ads** and **official non-political ads**. Note that this dataset does not contain a representative sample of political ads as they are ads received by people who answered ProPublica's call for action to install the tool.

AdAnalyst dataset Similar to the extension provided by ProPublica, AdAnalyst collects the ads users see on their Facebook timeline [3]. The AdAnalyst dataset contains over 500k ads from users in various countries. For this study, we keep only ads in English (detected using text-blob python library [90]) and that targeted users in the US between October 2018 and May 2020. For this, we use information about ad targeting available in the "Why am I seeing this ad?" button and select only ads targeted at people who live in the USA or visited places in the USA recently. The resulting dataset contains 9k unique ads (198 ads with "Paid for by" disclaimer and 8802 without). This dataset does not have votes from volunteers.

Ethical review board and reproducibility Both data collection by ProPublica and AdAnalyst were approved by the respective ethical review boards. The ProPublica data is available to the public through a request form [74]. The 9k ads from AdAnalyst, the data collected from the Prolific studies, and other supplemental material is available at `http://lig-membres.imag.fr/gogao/www21.html`.

4.2 Disagreement on political ads

The base to detect political ads reliably is to agree on which ads should be considered political and which ads should not. In this section we look at whether ad platforms, volunteers, and advertisers agree on which ads are political.

4.2.1 Disagreement across ad platforms

The Terms of Services of different ad platforms provide information on which ads they consider political. We review the definitions of online political advertising that were presented in sec. 3.

Overall there are three categories of political ads: **ads from or about a political figure or political party**, **ads about elections**, and **ads about social issues**. While the precise definition of political ads varies across ad platforms, the most significant difference is that Twitter and Google do not consider ads about social issues as political, while Facebook and TikTok do. While it is certainly a debatable question whether or not social issue ads should be regarded as political, the EU Code of Practice on Disinformation mentions both issue ads and political ads as sensitive content. Our results will show the importance of considering social issue ads as political and why they complicate political advertising regulations.

4.2.2 Disagreement among volunteers

At least three volunteers have labeled each ad in the ProPublica dataset as being political or non-political. The volunteers were given no instructions for what ads they should consider as political, and users were left to decide based on their instinct. To observe to which extent volunteers agree on what ads are political, Figure 4.1 plots the distribution of the number of political votes divided by the number of all votes for each ad in the ProPublica dataset. We denote this fraction as fr. A fraction $fr = 1$ means that everyone agrees that the ad is political, while a fraction $fr = 0$ means that everyone agrees that the ad is not political. The plot shows that for more than 50% of the ads, at least one volunteer disagrees with the others (fr is neither 0 nor 1), which shows that *deciding whether or not an ad is political is debatable for more than half of the cases.*

To distinguish ads on which users agree they are political from the rest, we split the ads into four disjoint *ad groups* based on the volunteer votes. We will analyze them separately. The groups are defined as follows:

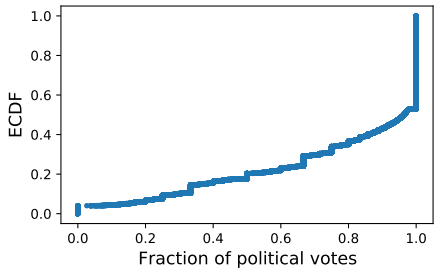


Figure 4.1: ECDF of the fraction of political votes for the ads in the ProPublica dataset.

Table 4.2: Number of ads in different ad groups (based on volunteer votes) and overlap with ads labeled as political by advertisers. † ProPublica was not able to give us access to ads that did not have at least one political vote and that were not labeled as official political ads.

	All	Official pol.	Official non-pol.
strong political ads	26k	96%	4%
political ads	19.7k	93%	7%
marginally political ads	7.6k	74%	26%
non-political ads	1.3k	100%	NA†

- **strong political ads**: ads with $fr = 1$, i.e., where everyone agrees that they are political;
- **political ads**: ads with $0.5 \leq fr < 1$, i.e., where there is some disagreement, but the majority labels them as political;
- **marginally political ads**: ads with $0 < fr < 0.5$, i.e., where there is some disagreement, but the majority labels them as non-political;
- **non-political ads**: ads with $fr = 0$, i.e., where everyone agrees that are non-political.

There are 26k **strong political ads**, 19.7k **political ads**, 7.6k **marginally political ads**, and 1.3k **non-political ads**.

4.2.3 Disagreement between volunteers and advertisers

The ProPublica dataset provides data on whether an ad was labeled as political by the advertiser itself (see Section 4.1). Table 4.2 presents the overlap between ads labeled as political by volunteers and ads labeled as political by advertisers (the **official political ads**). The table shows that 96% of **strong political ads**, and

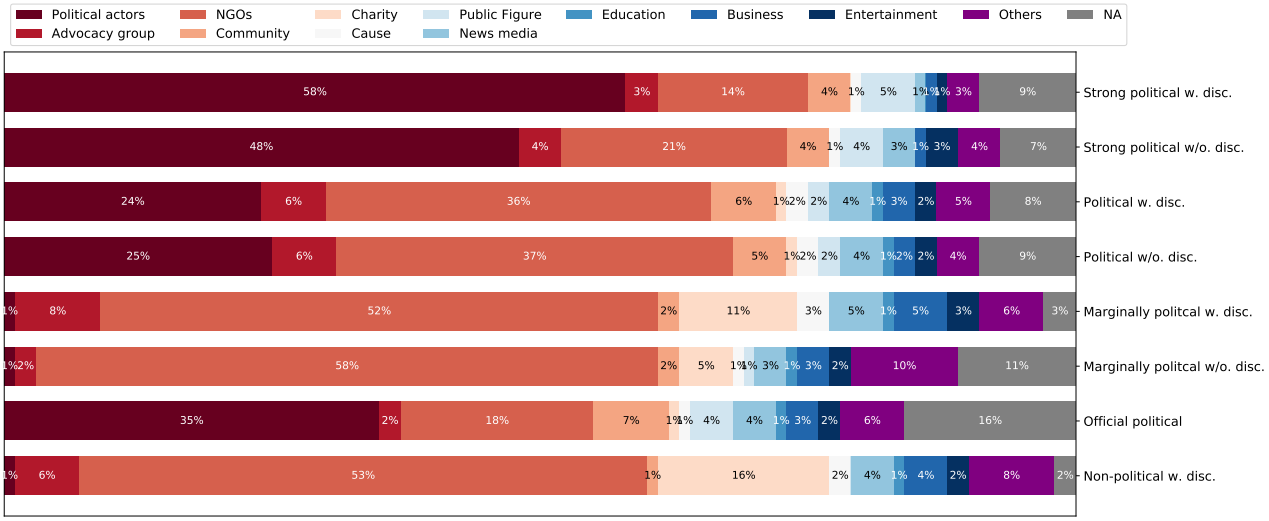


Figure 4.2: Breakdown of advertisers categories for different groups of ads for ads with and without "Paid for by" disclaimer.

93% of **political ads** were also declared as political by advertisers. Hence, most ads considered political by the majority of volunteers are also considered political by advertisers. There are, however, 4% of **strong political ads** and 7% of **political ads** that advertisers did not label as political.

The more surprising finding is that advertisers label as political a large majority (74%) of **marginally political ads**. Looking the other way around, 83% of **official political ads** are labeled as political by most volunteers. In contrast, 15% of **official political ads** are only labeled as political by a minority of volunteers, and 2% of **official political ads** are not labeled as political by *any* volunteer. Hence, many ads considered political by advertisers are not regarded as political by volunteers. While the reasons can be diverse (this is the subject of Section 4.4), we conclude that *there is currently a significant discrepancy between the ads labeled as political by advertisers and by volunteers.*

Takeaway: The assumption that we can clearly distinguish political from non-political ads does currently not hold as there are significant disagreements between ad platforms, volunteers, and advertisers on which ads are political. Therefore, it is problematic to apply restrictions on political ads if the decision of whether an ad is political depends on the person labeling it.

4.3 What gets labeled as political

This section provides a general view of ads labeled as political by volunteers and advertisers and analyzes who sends them and what are they talking about. This analysis is relevant for informing the debate on definitions of political ads and understanding the impact of potential regulations. The next section will focus on which ads lead to disagreement.

To interpret the results, we need to know the precise conditions in which the labeling happened. The ProPublica *volunteers* were given no instructions for what ads they should consider as political, and they were left to decide based on their subjective beliefs and background knowledge. However, volunteers could see if an ad was labeled as political by the advertiser itself (these ads have a "Paid for by" disclaimer on Facebook). We present results separately for ads that run with a disclaimer and ads that run without a disclaimer to isolate the potential effect of the "Paid for by" disclaimer.

Advertisers have to self-declare if they send political ads (as defined by Facebook's ToS). However, there is no public information on how Facebook enforces this policy [82]. Hence, ads labeled as political by advertisers are either a product of their own belief that their ad is political; or the result that Facebook constrained them to label the ad as political to run on the platform (maybe due to false positives in their enforcement algorithms).

4.3.1 Analysis of advertiser categories

To characterize advertisers we analyze their category. Advertisers need to create a Facebook Page and select from a pre-defined list a category for their page such as "Software Company" or "Political Party" [29]. We use the advertiser's ids available in the dataset to collect their category using the Facebook Graph API. Some pages no longer exist, we were able to extract categories for 6476 ProPublica advertisers (82%). Figure 4.2 plots the breakdown of the corresponding advertisers categories for **strong political ads**, **political ads**, **marginally political ads**, **official political ads** and **non-political ads**. We group similar advertiser categories:

Figure 4.2 shows that most **strong political ads** come from political actors (58% w. and 48% w/o. disc.), but a significant fraction of ads also come from NGOs (14% w. and 21% w/o. disc.), communities (4% w. and 4% w/o. disc.), and advocacy groups (3% w. and 4% w/o. disc.). In the **political ads** group, a smaller fraction of ads come from political actors (24% w. and 25% w/o. disc.), much more from NGOs

Table 4.3: Examples of ads from advertisers with different categories.

Advertiser	Text	*fr*	disc.
Category: Cause			
UnRestrict Minnesota	96% of Minnesotans don't know the abortion laws in our state.	1	w.
Care2	U.S. Wildlife Services is putting the safety of people: animals at risk in its attempt to control wild predators. Tell them to STOP using taxpayers money to kill wild animals lethally!	0.75	w.
Oregon Forests Forever	Brave men and women from Oregon are helping to fight fires in California	0.37	w.
Home Ownership Matters	Do you want Congress to invest in infrastructure? Click here to sign the petition.	0.33	w.
Category: Charity			
USA for UNHCR	Should America turn away from this child? Not now, not ever. It's not who we are.	1	w.
World Food Programme	I call on warring parties to allow the constant flow of food for innocent and starving people in Yemen. Add your voice to our petition today.	0.66	w/o.
ChildFund International	She wants a childhood free of worry and a future full of promise.	0	w.
USA for UNHCR	All donations MATCHED for a limited time. People in Syria are still fleeing for their lives. UNHCR needs your help to provide the shelter, food and medicine they need to survive.	0.33	w/o.
Category: Community			
Yes for Washington Elementary Students	Vote YES on the WESD Override to protect full-day kindergarten, music, art, and physical education in our schools.	1	w.
North Carolina Citizens	We have a new survey for North Carolina. Please click the link below to share your thoughts	0.8	w/o.
Healthy Me PA	Workplace violence is 4x more common in the health care industry. Here's how you can help:	0.3	w.
Protect Coyote Valley	Time and time again, threats of development have been made in Coyote Valley, with some succeeding. We want to see Coyote Valley permanently protected for our wildlife and for our children. All we need is your signature	0.4	w.
Category: Business			
Dissent Pins	Stand for democracy on election day and every day with our Count Every Vote pin.	1	w.
Ben and Jerry's	Vote YES on 4 and reinstate voting rights for 1.4 million Floridians!	0.96	w.
CREDO Mobile	Help us decide how to allocate our $50k donation to 5 progressive environmental organizations fighting for climate justice.	0.33	w.
Steady Returns, LLC	Everyone deserves great financial advice!	0.2	w/o.
Category: NGOs			
Democratic Attorneys General Association	Now that we know Joe Biden will be the nominee, we want to know who you think he should pick as his V.P.? Hurry, this round closes soon and we are still missing your response.	1	w.
Pennsylvania Spotlight	Voting from home is easy. By taking thirty seconds to request a ballot, you can fill your ballot out on your couch and mail it in.	0.63	w.
National Audubon Society	Birds and their habitats are under attack, but with your help we can fight back. This Earth Day your monthly gift will go twice as far to protect birds and the places they need..	0.25	w.
FOUR PAWS International	Stray animals are starving in India, will you give them your much-needed support?	0.33	w/o.
Category: Political actors			
Arati Kreibich for Congress	Republicans are suppressing the vote through mass voter purges, polling place closures, and burdensome voter ID laws. Tell the Senate: restore the Voting Rights Act!	1	w.
Bernie Sanders	We are about to make history and I want you to be a part of it. Our campaign is trying to reach 1 million campaign donors faster than any campaign in American politics, and we are VERY close. Can you make a contribution right now to become one of our first million donors?	0.66	w.
Tina Smith	Meet Senator Tina Smith: a big fan of dogs, donuts, and Minnesotans.	0.66	w/o.
Judge Brian Hagedorn	Click here to hear how an adopted daughter changed the Hagedorn family!	0.33	w.

(36% w. and 37% w/o. disc.), and we also see more ads from advocacy groups (6% w. and 6% w/o. disc.), news media (4% w. and 4% w/o. disc.), and communities (6% w. and 5% w/o. disc.). In the **marginally political ads** group, only (1% w. and 1% w/o. disc.) of ads come from political actors, the majority (52% w. and 58% w/o. disc.) from NGOs and charity organizations (11% w. and 5% w/o. disc.), some ads come from news media (5% w. and 3% w/o. disc.) and businesses (5% w. and 3% w/o. disc.). In the **official political ads** group, we see a similar diversity in the advertisers labeling their ads as political. *Many countries' specific electoral legislation only regulate (and impose restrictions on) ads from political actors. However, we see that there is a wide range of advertisers pushing political ads online and that volunteers do label ads from these advertisers as political; hence, prompting for updating legislation.*

Facebook is explicitly exempting news organizations from labeling their ads as political even if they are about political issues [32]; however, yet do seem to consider these ads as political. *This raises the question of whether ads from news media should be treated as political ads. On one side, political journalism is different from political propaganda; on the other side, news media has been used as a tool to manipulate users, and many unauthentic news aggregators are emerging with the purpose of promoting a political agenda [10].*

Table 4.3 presents examples of political ads from different categories of advertisers such as community, NGO, or business. For each ad, the table shows the fraction of political votes divided by all votes from volunteers and whether the ad was labeled as political by the advertiser itself. The table shows that there is a wide diversity of ads getting labeled as political. For instance, we can see an ad from the ice-cream company "Ben and Jerry" (a business) that is inciting citizens to vote, and an ad from the "Democratic Attorneys General Association" (an NGO) that is asking people who should be the V.P. of Joe Biden. Such ads have a clear association with elections. In the table, we also see many ads, such as the ones from the "World Food Programme" and the "USA for UNHCR" (Charities), that address social issues but do not seem to have any evident association to elections or legislation. *The critical point to recognize is that ads labeled as political can have a very different level of "politicalness", going from straight advocacy messages addressing abortion issues to ads merely asking for a donation to end world hunger.*

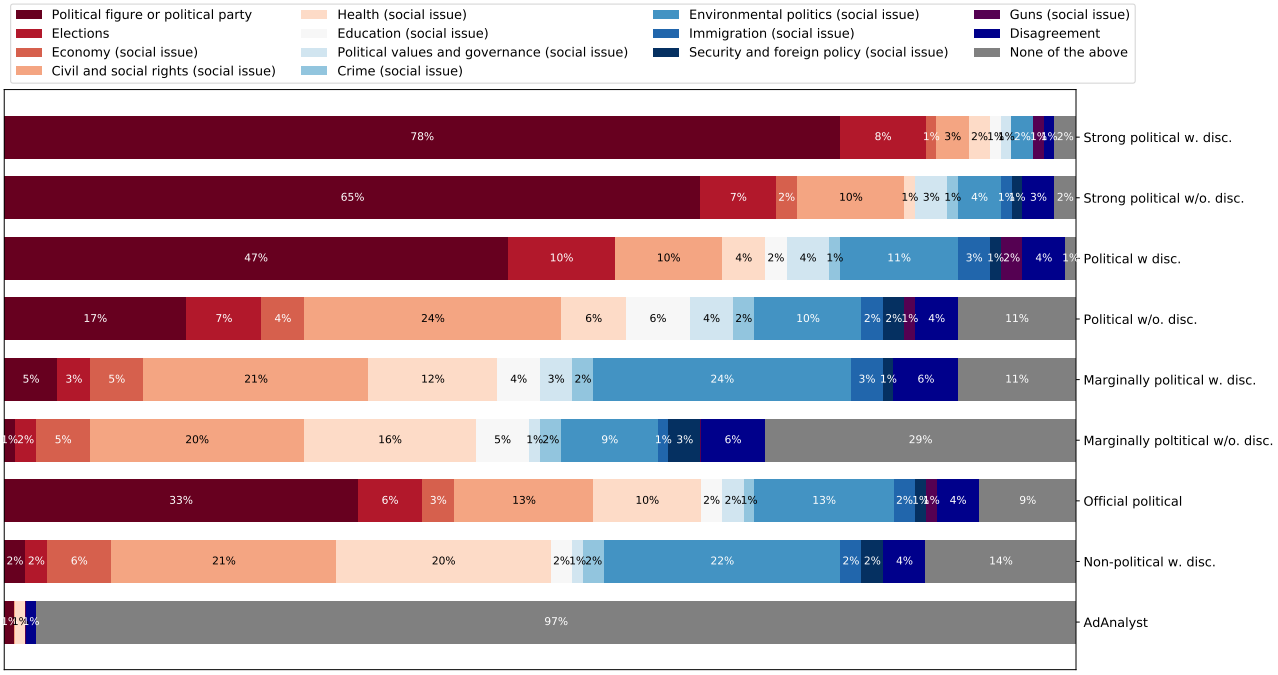


Figure 4.3: Breakdown of the political and social issues discussed in ads for the different groups of ads with and without disclaimer.

4.3.2 Analysis of ad messages

To gather grounded information about the topics of ads labeled as political, we took a random sample of 300 ads with a "Paid for by" disclaimer and 300 ads without "Paid for by" disclaimer from each **strong political ads**, **political ads**, **marginally political ads**, 300 ads **non-political ads**, and 200 ads without disclaimer from AdAnalyst. While we picked both ads with and without a disclaimer, we did not show the disclaimer in our surveys. We set up a survey on Qualtrics [76] where for each ad, we ask respondents questions about the ad's message. We hired workers through Prolific [73], and we redirected them to fill out the survey on Qualtrics. Each worker had to label 20 random ads from the pool of 2300 ads, and each ad was labeled by three workers. We selected workers that are residing in the USA since all the ads used in the experiments targeted people who lived in or visited the USA. The median amount of time that workers spent on the survey was 12 minutes.

Each survey had an instructions page, followed by 20 pages each containing one ad to label. For each ad, we asked the following questions:

(1) "Is this ad made by, on behalf of, or about a political actor? (such as a candidate for public office, a political figure, a political party or advocates for the outcome of an election to public office)"; (2) "Is this ad about elections? (such as

referendum or ballot initiative, including "go out and vote" or election campaigns)"; and (3) "Does this ad refer to a social issue? (such as civil and social rights, ...)". Workers were allowed to answer yes to all the questions. If workers selected that the ad is about a social issue, we asked them which social issue: "Which social issue is this ad talking about?" Workers had to choose from the following list: civil and social rights, crime, economy, education, environmental politics, guns, health, immigration, political values and governance, security and foreign policy. We considered these social issues because they appear in the Facebook definition of political ads [28]. Workers were allowed to select multiple social issues if needed.

If workers answered no for all three initial questions (the ad is not about a political figure, election, or social issue), they were asked to choose from a list "What topic describes best the ad". We took the list of 23 topics from the Interactive Advertising Bureau (IAB) categories [31]. Note that we did not ask workers whether the ad is political or not; we just asked them questions about its message. Figure 4.3 shows the breakdown of the political or social issue discussed in an ad according to Prolific workers for different ad groups for ads with and without disclaimer. For each ad, we pick the ad topic chosen by the majority of workers or mark it as disagreement if no two workers chose the same ad topic or if two topics had an equal number of votes. We attributed all ads about both a political figure and a social issue or a political figure and election to the political figure group, and all ads about both an election and a social issue to the election group. For clarity, all ads for which the majority of workers chose a (non-political) IAB topic are marked as "None of the above" in Figure 4.3.

Figure 4.3 shows that all groups of ads contain most of the ad topics we consider. We see higher fractions of about a political figure or political party and ads about an election in the **strong political ads** (78%+8% w. and 65%+7% w/o. disc.) and higher fractions of social issues ads in the **political ads** (38% w. and 61% w/o. disc.) and **marginally political ads** (75% w. and 62% w/o. disc.). In the **official political ads** group, there is also a high fraction (48%) of social issue ads. The non-political AdAnalyst ads are shown as control. Indeed less than 2% of these ads are labeled as being about a political figure, election or social issue. *Social issue ads are only considered political by Facebook and TikTok, not by Google and Twitter. However, these results tell us that a large proportion of the ads volunteers and advertisers label as political are about social issues. Hence, it is crucial to consider social issue ads as political as well.*

Figure 4.3 shows that some ads (2% w. and 2% w/o. disc. of **strong political ads** and 1% w. and 11% w/o. disc. of **political ads**) were not labeled by workers as being about a social issue, a political figure or election. Since there is no expert ground truth, we cannot say whether labels from volunteers or labels from workers are better. Nevertheless, the (non-political) IAB topics that were mentioned the most by workers were society, health & fitness, education and science. *This raises questions on where to drawl the line between ads about civil and social right and ads about society; or ads about health as a social issue and ads about health & fitness as a lifestyle.*

One might decide that **marginally political ads** should not be treated as political because only a minority of volunteers labeled them as political. Figure 4.3 shows that 5%+3% w. disc. and 1%+2% w/o. disc. of **marginally political ads** do contain ads from a political figure or political party or elections. In addition, 21% w. disc and 20% w/o. disc. ads are about civil and social rights, and 24% w. disc. and 9% w/o. disc. are about environmental politics. The numbers look similar for **non-political ads**. Marginally political ads do contain a significant number of political ads as defined by the Facebook ToS. *These results show that **marginally political ads** should not be ignored because they might contain ads about social issue and ads where only a few people have the right background knowledge to detect them as political.*

Takeaway: Our results show that a large fraction of ads labeled as political are about social issues and do not mention a political actor or elections. Hence, it is crucial to consider ads about social issues as political. Our results also show that a wide range of ads are getting labeled as ads about social issues. Hence, since many legislative projects are considering to severely restrict micro-targeting [54] or ban such ads altogether; we need to decide whether we want ads (with no apparent link to elections and legislation) coming from charities or communities to be subject to the same restrictions as ads that advocate polarizing issues. Such restrictions could hurt a wide range of humanitarian civil organizations.

4.4 Learning from disagreement

The previous section showed that a very diverse set of ads get labeled as political. This section analyzes the ads that lead to disagreement among volunteers and between volunteers and advertisers. This analysis is relevant for refining political ads' definition and improving the processes and instructions for labeling ads.

4.4.1 Volunteers vs. advertisers

To understand why advertisers and volunteers disagree on ads being political, we examine separately ads that seem to be underreported by advertisers and ads that seem to be underreported by volunteers.

Ads underreported by advertisers These are the **strong political ads** and **political ads** without disclaimer. Table 4.2 shows that 4% of the **strong political ads** and 7% of the **political ads** are not labeled as political by advertisers. There are several possible (non-exhaustive) explanations: (1) advertisers do not comply with the ToS (e.g., they willingly do not label their ads as political to avoid scrutiny), i.e., volunteers are right; (2) advertisers underreport certain categories of political ads, i.e., advertisers and volunteers have different interpretations of which ads are political; and (3) volunteers misinterpret the ads' message, i.e., advertisers are right.

Figure 4.2 presents the breakdown of advertiser categories and Figure 4.3 the breakdown of ad types corresponding to **strong political ads**, and **political ads** without disclaimer. A significant fraction of advertisers are political figures (48% in **strong political ads** and 25% in **political ads**), and a significant proportion of ads refer to a political figure or political party and elections (65%+7% for **strong political ads** and 17%+7% for **political ads**). *Hence, more than half of **strong political ads** and **political ads** without disclaimers are not compliant with Facebook's ToS. These results confirm previous findings that advertisers sometime do not label their ads as political and the need for better enforcement mechanisms [82].*

A large fraction of ads without a disclaimer (23% of **strong political ads** and 61% of **political ads**) are about social issues. Recall that we excluded from this category ads labeled as social issues but mentioning a political figure or elections. Tables 4.6 and 4.7 show some examples of ads about civil and social rights and environmental politics in **strong political ads** and **political ads** without disclaimer. These ads are on topics such as climate change and healthcare, which are very politicized issues in the US, and give valid reasons to volunteers to label them as political.

To understand whether ads about some social issues are less disclosed by advertisers than others, for each ad topic, we compute the fraction of ads that do not have a disclaimer in the **strong political ads** and **political ads** groups. Ads about economy (0.15), civil and social rights (0.28), and security and foreign policy (0.27) have the lowest fraction of ads with a disclaimer. In contrast, ads about political figures (0.6), election (0.57), and environmental politics (0.49) have the highest fractions of

ads with a disclaimer. *This shows that advertisers are underreporting ads about social issues, especially if they are about economy or civil and social rights.*

For 2% of **strong political ads**, and 11% of **political ads** w/o. disc. workers did not label them as being about a political figure, election, or social issue; which means that no one besides volunteers labeled them as political. Table 4.5 shows a few examples of such ads. These ads seem to address some issues but are not clearly related to the social issues provided to workers. This raises an interesting dilemma: if someone labels an ad as political (without being forced or by mistake), can they be wrong?

Ads underreported by volunteers These are **non-political ads** and **marginally political ads** with disclaimer. There are 1.3k **non-political ads**, and 5.6k **marginally political ads** (74%) labeled as political by advertisers. There are various reasons why advertisers would label their ads as political while all/most volunteers labeled them as non-political: (1) advertisers might be *forced* to label ads as political (even if they are not) because of false positives in the enforcement mechanisms implemented by the ad platform; (2) advertisers might think that disclaimers would bring more attention to their page; (3) advertisers understand better why their ads should be political, and volunteers underreport such ads; etc. Figure 4.3 shows that a significant fraction (14%) of **non-political ads** are labeled as not being related to a political figure, election or social issue by workers; meaning that no one besides advertisers are considering these ads as political. Table 4.5 shows a few examples of such ads. Indeed, the majority of these ads do not seem to be political. *Since substantial restrictions are envisioned for political ads, it is essential to know what enforcement mechanisms are put in place by ad platforms to understand what is the impact of false positives in their algorithms.* Non-political ads mislabeled as political is also problematic when building detection methods that use political ads labeled by advertisers to train models. Thus, it is important to look for poisoning attacks when building such models.

Figure 4.3 shows that the majority of **non-political ads** and **marginally political ads** without disclaimer are related to civil and social rights (21% and 20%), health (20% and 16%) and environmental politics (22% and 9%), while only a few refer to political actors (2% and 1%) or elections (2% and 2%). Figure 4.2 shows that these ads come mostly from NGOs (53% and 58%), news media (4% and 3%), businesses (4% and 3%), and charities (16% and 5%), while only a few (1% and 1%) come from political actors. *Hence, it seems that volunteers underreport many ads about a social issue, especially about civil and social rights and health, and ads from advertisers such*

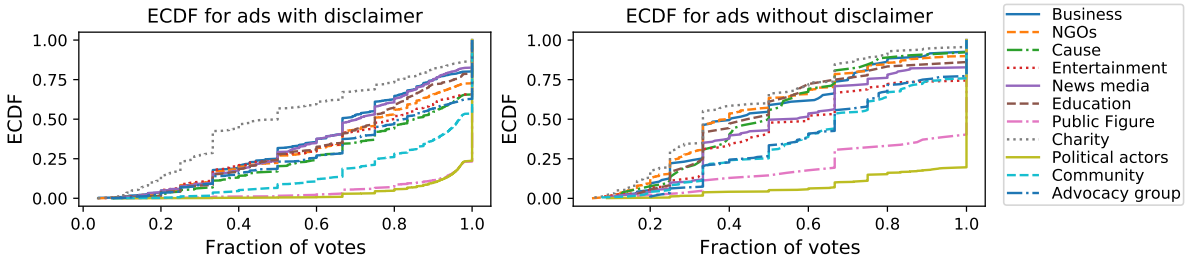


Figure 4.4: ECDF of the fraction of political votes for ads from different advertiser's categories in **strong political ads**, **political ads**, and **marginally political ads**.

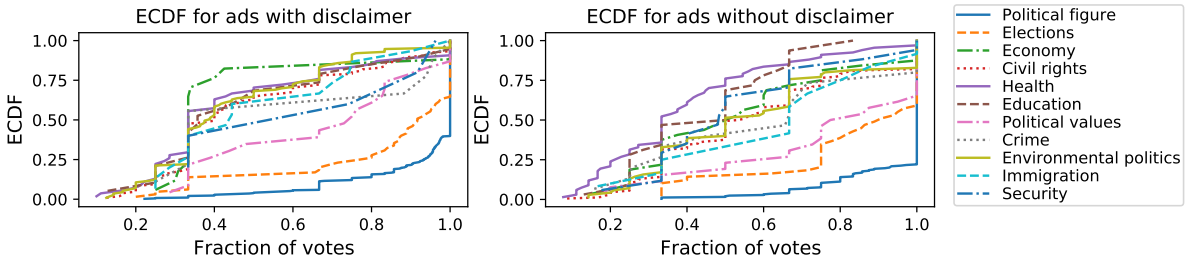


Figure 4.5: ECDF of the fraction of political votes for ads with different ad topic in **strong political ads**, **political ads**, and **marginally political ads**.

NGOs, and charities. Table 4.6 and 4.7 present examples of ads about civil and social rights and environmental politics in **non-political ads** and **marginally political ads**. We see that most of these ads are related to social issues, but volunteers might not consider them as political because there is no apparent association with elections or legislation.

Takeaway: Two main factors contribute to disagreement between advertisers and volunteers: (1) advertisers mislabel ads as political or non-political (maybe to avoid scrutiny; maybe because they are forced to label their ads as political by enforcement mechanisms put in place by ad platforms); and (2) both advertisers and volunteers underreport ads about social issues. Part of the problem may be that the definition of ads about social issues is too broad which leads to different interpretations among people. This raises the question of whether *all* ads related to social issues should be considered political, and if not, how should we filter social issue ads that are not political. For example, one possibility would be to consider as political only ads about social issues that could directly or indirectly impact elections or legislation or that address polarizing issues.

4.4.2 Volunteers vs. volunteers

To investigate which ads lead to disagreement among volunteers, we check if there is more disagreement on ads coming from specific advertisers and ads about particular political or social issues.

To see if ads from certain categories of advertisers lead to more disagreement, for each advertiser category, we group all corresponding ads (from **strong political ads**, **political ads**, and **marginally political ads**). Figure 4.4 shows the ECDF of *fr* for each group. Ads with a *fr* close to 0.5 have the highest level of disagreement (half of the volunteers label them as political and half as non-political). We split the analysis on ads with disclaimer and ads without a disclaimer since the disclaimer might have impacted how volunteers voted. We see in Figure 4.4 that the distributions shift to the right (more political votes) when ads have a "Paid for by" disclaimer. However, we cannot attribute this shift solely to the presence of disclaimers because ads with disclaimers might also have messages that are "more political". The plot shows that at least 10% of ads in each advertiser category has $fr = 1$, which means that at least *some* volunteers are not bothered by the fact that the ad is coming from non-traditional political actors. Figure 4.4 shows that ads coming from political actors and public figures achieve the highest agreement, 85% have $fr = 1$. Besides, ads from communities and advocacy groups tend to be seen as more political, while ads from charities as less political. Ads from other advertisers such as NGOs, causes, news media, education, and businesses are somewhere in between, leading to the highest level of disagreement. To get definite proof if the advertiser category influences the decision (and not the message of the ad), we would need a conjoint analysis that tests the same ad message with different advertisers but our data does not permit such analysis. In any case, *platforms and policymakers should clarify how much consideration should be given to the advertiser when labeling ads as political.*

To see if ads from certain ad topics lead to more disagreement, for each ad type, we group all corresponding ads (from the 1800 ads labeled by Prolific workers in **strong political ads**, **political ads**, and **marginally political ads**). Figure 4.5 shows the ECDF of *fr* for each group (we only show groups for which we have more than 20 ads labeled). We can see that the highest agreement is among ads that mention political figures and elections, while, as expected, the highest disagreement is on various social issue ads. We performed a pairwise Kolmogorov-Smirnov statistical tests between the distributions. Ads about elections and political figures are statistically different than the rest; but most of the social issue ads are not statistically different between them. To see why for a particular social issues, some ads have higher *fr* than others,

Tables 4.6-4.7 show examples of civil and social rights ads and environmental politics ads for different ad groups. We see that the ads address a wide range of topics (e.g., abortion, wildlife, hunger), they call for various actions (e.g., sign petitions, surveys, donations, call an elected representative) and try to provoke various sentiments (e.g., pride, anger, fear). Ads that address climate change and pollution are seen as more political, while ads about wildlife protection are seen as less political. Besides, ads that refer to problems in the U.S. (ad from NRDC) are seen as more political than ads that refer to problems in other countries (ad from Care2). While these are only anecdotal examples, they emphasize the complexity of deciding which ads are political.

Limitation: There are other reasons for disagreement that we could not analyze with this dataset. For example, the background knowledge of volunteers might impact how they vote (the political nuance of an ad is only recognized by some) or the political ideology of volunteers impacts how they vote. These questions are essential for recruiting moderators, and we leave them for future work.

Takeaway: Ads from NGOs, causes, news media, education, and businesses and ads on social issues lead to the highest disagreement among volunteers. To distinguish better political from non-political ads, we would need policy recommendations that clarify the perimeter of social issue ads. This raise a multitude of questions such as: Should we treat ads about more politicized issues differently than ads about less politicized issues? Should we treat social issues depending on the country? Should we treat ads that call for precise actions differently than ads that just inform citizens? Should we define social issues at a smaller granularity (in both topics and locality) than currently? How should the system adapt to emerging social issues? How much weight should be given to the advertiser's identity (as opposed to just the ad content)?

4.5 Classification and disagreement

Traditional supervised classification algorithms create models from positive and negative examples that we feed in the training phase. The previous sections showed significant discrepancies between ads labeled as political by advertisers and ads labeled as political by volunteers. Hence, this raises the question of whether classifiers trained on one or the other would result in significantly different models. Intuitively, if the training examples are biased, the models will be different, while if the training examples are representative of political ads in general, the resulting models will be

similar. This section investigates how discrepancies in positive labels from advertisers and volunteers impact the resulting classification models.

For the evaluation we split the ProPublica dataset in two equal-size slices of 28k ads: S_1 and S_2. We use S_1 as the training and validation dataset and S_2 as the holdout/test dataset. We build four models using four different sets of positive labels but the same negative labels. As negative examples, we took 7.5k ads in English from AdAnalyst without the "Paid for by" disclaimer (see Section 4.1).

The M_{op} model: the positive labels are a random sample of 8000 **official political ads** from S_1. M_{op} is trained with positive examples from *advertisers*.

The M_{sp} model: the positive labels are a random sample of 8000 **strong political ads** from S_1. M_{sp} is trained with only positive examples where all volunteers agree they are political, $fr = 1$.

The M_{mp} model: the positive labels are a random sample of 4000 **political ads** and 4000 **strong political ads** from S_1. M_{mp} is trained with positive examples where the majority of volunteers consider the ads as political, $fr > 0.5$.

The M_{1p} model: the positive labels are a random sample of 2600 **strong political ads**, 2600 of **political ads** and 2600 **marginally political ads** from S_1. M_{1p} takes as positive examples all ads where there exist at least one user that labeled it as political, $fr > 0$.

To build the different models, we used Naive Bayes. While Naive Bayes is neither new nor sophisticated, it was shown by [82] that it achieves high accuracy for detecting political ads and outperforms other methods. The classifiers only take as input the ad's text, and as pre-processing, we deleted all Html tags, stop words, and punctuation. We used Count Vectorizer for text embedding [80].

We performed 10-fold cross-validation for each classifier over its specific training-validation dataset that contains 8000 positive and 7.5k negative examples. Table 4.8 presents the average accuracy and true positive rate for a 1% false positive rate for the four classifiers. For systems where the fraction of positive examples (political ads) is much smaller than the fraction of negative examples (non-political ads), it is essential to limit the rate of false positives (non-political ads labeled as political); hence, we are interested in true positive rates for a 1% false positive rate. The table shows that all classifiers achieve high accuracy of over 95%, but only M_{op}, M_{sp}, and M_{mp} achieve true positive rates of more than 90%. The lower true positive rate of

M_{1p} (85%) is expected as it has a more challenging task because it is trained and tested with more debatable political ads.

The main challenge in evaluating the classifiers is that we do not have a gold standard collection of political and non political ads. Table 4.8 only tells us *how good the models are at identifying the same kind of political ads with the ones they were trained on, but not how good they are at identifying political ads in general.* Hence, we look next at how these models perform on detecting other kinds of political ads then those they were trained on.

We use the four models to make predictions for all ads in S_2. To predict that an ad is political, we took the threshold corresponding to a 1% FRP for each classifier. Table 4.9 shows how well the four models are at identifying **official political ads**, **strong political ads**, **political ads**, **marginally political ads**, and **non-political ads** in S_2. As negative examples, we used 1000 ads from AdAnalyst that do not have a disclaimer and were not used for training.

Table 4.9 shows that M_{sp} has the lowest number of false positives, while M_{op} has the largest number. For detecting **strong political ads**, all models detect more than 94% of ads. For detecting **political ads** the M_{op} and M_{mp} perform the best (detecting over 94% of ads). For **marginally political ads**, M_{1p} and M_{mp}, perform well (over 86% detection), while M_{1p} has a 84% detection. For **non-political ads**, M_{op} and M_{mp} label more than 85% as political.

The detection rates of M_{op} and M_{mp} are similar across different datasets, with M_{mp} performing better especially on **marginally political ads** and **non-political ads**. Hypothetically if the resulting classifiers would label as political the precise same ads, the input data is representative of political ads, and who is labeling the training data (be it advertisers or volunteers) does not matter. To understand whether M_{op} and M_{mp} label the same ads as political, we computed the fraction of ads labeled by both models as political over all ads for different ad groups in S_2. The data shows that the two models have an overlap of 97% in **strong political ads**, 94% in **political ads**, 83% in **marginally political ads**, and 82% in **non-political ads**. *These results show that the overlap is relatively high, but discrepancies in the input data do transfer to discrepancies in the output data. Hence, we need to consider how biases in labeling are impacting classification results and whether this may lead to unfairness against certain advertisers.*

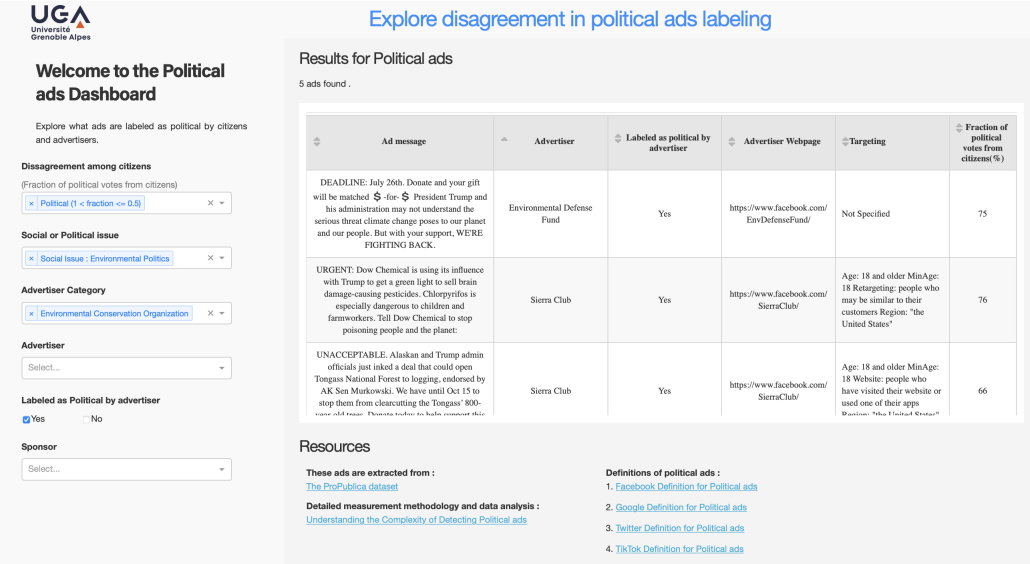


Figure 4.6: The interface of the server to explore disagreement in online political advertisements.

4.6 Service to explore disagreement

We created a service to explore disagreement in online political advertising. For the implementation, was choosing Python language and its original framework Dash [71]. Dash is a framework for building interactive web applications. It is specifically designed for creating data-driven web applications that require data visualization and user interaction. With the help of the service, users are able to conduct comprehensive searches of political ads based on various parameters such as the level of disagreement, topics, advertisers, advertisers' categories, and sponsors. Fig.4.6 shows the interface of our service.

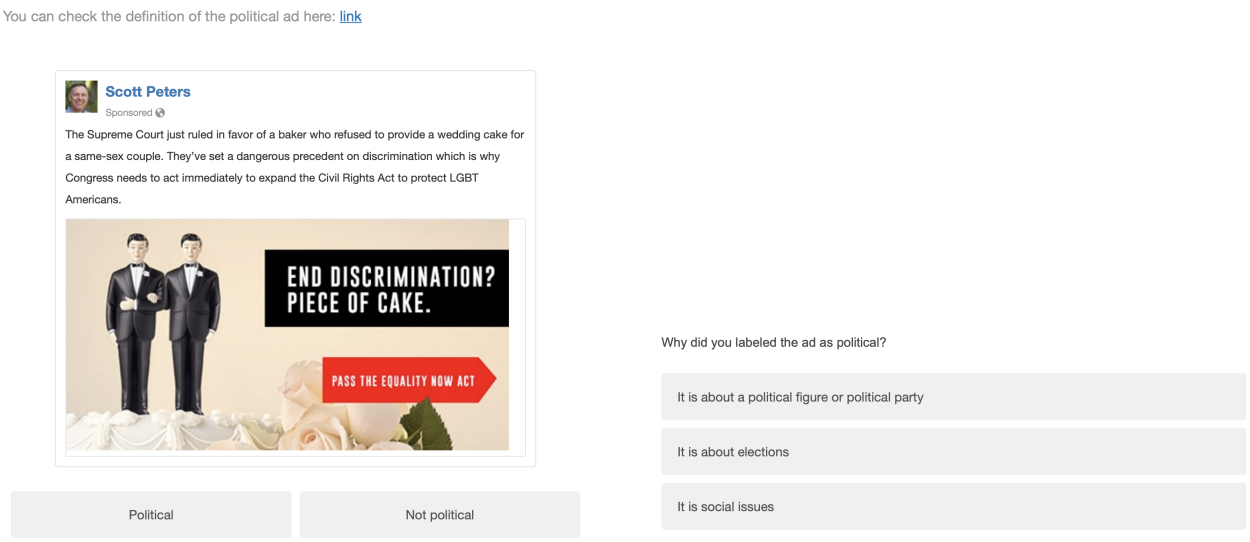
The service is available at: `https://facebookads.imag.fr`

4.7 Impact of task design on ad labeling

This section investigates what effect has the ad labeling process on what ads humans label as political. We focus on two questions:

Q_1: *How does the definition of what is a political ad impact ad labeling?*

Q_2: *How does social influence impact ad labeling?*



(a) (b)

Figure 4.7: Examples of surveys' questions.

4.7.1 Experiment design

To answer these questions we setup three experiments on Prolific where we asked workers to label ads as political or not. Each survey had an instructions page, followed by 20 pages that contained one ad to label (see Figure 4.7):

Exp_1: In the first experiment we did not provide any definition. The instruction page only contained the message: *In this survey you will be asked to decide whether an ad is* ***political*** *or* ***non-political****.*

Exp_2: In the second experiment we gave in the instruction page he definition of political ads from Facebook and we ask workers to label ads accordingly. We included a link to the definition in every page of the survey. In addition, for each ad, we asked workers to motivate their choice by selecting from the list of reasons (presented in Section 4.3.2) that were extracted from the Facebook definition for political ads. To match the quantity of work, we also asked workers to choose from a list of topics (taken from the IAB categories [31]) when they were labeling an ad as non-political.

Exp_3: In the third experiment, we did not provide a definition, but, for each ad we told workers what fraction of people labeled the ad as political previously (we used data from the Exp_1).

To create the surveys we used Qualtrics [76]. To test the influence of the ad labeling process on a wide range of ads we took a random sample of 200 ads from

strong political ads, 200 ads from **political ads** (100 with paid_for_by disclaimer and 100 without disclaimer) and 200 ads from **marginally political ads**. In addition, we took 200 random ads without disclaimer from AdAnalyst USA users. Note that, we do not show whether the ad has a disclaimer or not in our surveys. We use the same ads in the tree experiments. For every ad in every experiments we collected three labels. Every participant was given a random sample of 20 ads to label from the 800. Workers were only able to participate in one experiment.

4.7.2 Analysis of labels

To check if the impact on votes of our three experimental designs is statistically significant, we test the following two null hypothesis:

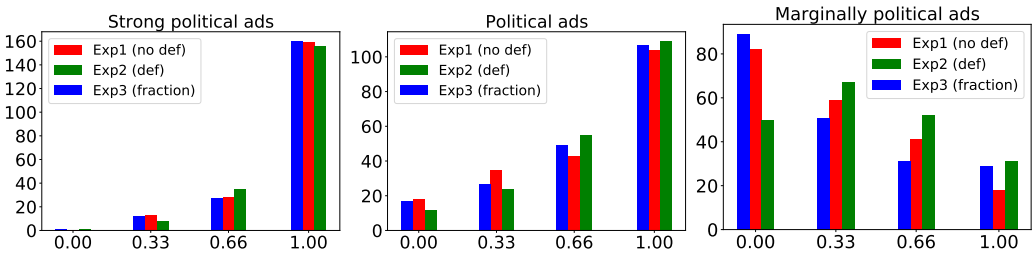
NH_1: The definition of political ads does not impact labeling.

NH_2: Social influence does not impact labeling.

To test the two hypothesis we check whether the distributions of the fraction of political votes for Exp_1 and Exp_2 are identical (NH_1) and if the distributions for Exp_1 and Exp_3 are identical (NH_2). To choose a proper statistical significance test we use the Kolmogorov-Smirnov test to check whether votes from the three experiments are normally distributed or not. The distributions are not normal. Hence, we use the Mann-Whitney U test which is a non parametric test to compare two independent distributions as our statistically significance test. The null hypothesis (the distributions are identical) is rejected if p_{value} is less than 0.05. The results show that we cannot reject NH_1 because it has $p_{value} = 0.059$ and we cannot reject NH_2 because it has a $p_{value} = 0.39$.

We further investigate whether there is a statistically signifiant impact on specific groups of ads: **strong political ads**, **political ads**, **marginally political ads** and non-political ads. We use the Wilcoxon T-test to check, inside each experiment, if votes on **strong political ads**, **political ads**, **marginally political ads**, and non-political ads come from the same distribution or not. We compare the four samples of data with each other. The results show that the samples of distributions are different from each other, so it is good to analyze them separately. We proved that samples of distributions are not normal distribution by using the Kolmogorov-Smirnov test.

Table 4.10 shows the p_{value} for the different groups of political ads between different experiments. The null hypothesis is rejected only for **marginally political ads** and NH_1. This shows that even if the definition for political ads did not have a significant



(a) **strong political ads** (b) **political ads** (c) **marginally political ads**

Figure 4.8: Distribution of the fraction of political votes for different groups of ads and different experiments.

impact on **strong political ads** and **political ads** ads it did have a strong impact on **marginally political ads**. This is good because these are the ads that are the most confusing and problematic. Contrary, NH_2 is not rejected for any group of ads, hence providing information about how other people labelled ads did not have an effect on votes.

Figure 4.8 plots distributions of the fraction of political votes for different groups of ads. The number of political votes increase in all presented groups of ads in the Exp_2, and this difference the highest in case of **marginally political ads**. Hence, providing the Facebook definition determined workers to see more political content in ads, however, the agreement does not seem to increase. This means there is room for improvement on both political ad definition and the instrumentation of the ad labeling process.

Discussion From an ad moderation perspective, consensus is desirable because the decision of what is a political ad is clear and is straightforward to exploit the labels to build automated ML algorithms to detect political ads. Consensus alone is however not enough, for example if you ask French people to label US political ads, they might only label ads from Donald Trump but nothing else because they are not familiar with US politics. French volunteers will have a perfect consensus, but they will miss on a lot of interesting political ads. Hence, a second important dimension is the **diversity** of ads labeled. Ideally we want to throw a large nest that is able to capture clear political ads coming from presidential candidates, but we would also want to capture ads coming from local candidates, and new social issues that arise in particular groups of people that are not well known nationally. Hence, ad labeling guidelines should attempt to increase consensus but not at the detriment of diversity.

4.8 Summary

Many agree that online advertising especially political adverting needs to urgently be regulated, but one missing key is how to reliably detect political ads. We attempt to dissect some of the complexity of labeling political ads. To our knowledge, this is the first study to show how ordinary people label ads as political, why they disagree and what are the implications for policymaking and enforcement algorithms.

We show that volunteers seem to underreport ads from NGOs, and charities (that are considered political by advertisers) and advertisers seem to underreport social issue ads (that are considered as political by volunteers). While disagreement can be alleviated through better guidelines to a certain degree, many ads addressing societal and humanitarian issues are intrinsically hard to label. We believe that the community needs a gold standard collection for political ads and to better define the perimeter of social issue ads. We hope our analysis can help policymakers and ad platforms to refine the definitions of political ads and their regulation.

Table 4.4: Categories that were represented by the same name later in the chapter.

Name	Categories
Business	Product/Service, Pet Supplies, Household Supplies, Software Company, Apparel & Clothing, Lawyer & Law Firm, Jewelry/Watches, Insurance Company, Retail Company, Clothing (Brand), Energy Company, Shopping & RetailSocial Service, Insurance Broker, Environmental Service, Advertising/Marketing, Outdoor & Sporting Goods Company, Wholesale & Supply Store, Financial Service, Business Service, Travel Company, Tax Preparation Service, Labor & Employment Lawyer, Tourist Information Center, Nutritionist, Mental Health Service, Pregnancy Care Center, Beauty, Cosmetic & Personal Care, Solar Energy Company, Internet Company, Solar Energy Service, Brand, Criminal Lawyer, Information Technology Company, Coffee Shop, Local Business, Consulting Agency, Food & Beverage Company, Gift Shop, Sunglasses & Eyewear Store, Public Relations Agency, Emergency Rescue Service
NGOs	Nonprofit Organization, Non-Governmental Organization (NGO), Organization, Religious Organization
Cause	Cause
Entertainment	Movie, Games/Toys, TV Show, Arts & Entertainment, Museum, Magazine, Broadcasting & Media Production Company, Author, Bookstore, State Park, Zoo, Cultural Center, Podcast, Festival, Show, Performance Art, Video Creator, Entertainment Website, Radio Station, Performance Art Theatre, Science Museum, Musician/Band
News media	Media/News Company, Media, TV Channel, TV Network, News & Media Website, Publisher, Newspaper, News Personality
Education	Educational Research Center, College & University, Education Website, Public School, Education
Public Figure	Public Figure
Charity	Charity Organization
Political actors	Government Organization, Political Organization, Political Candidate, Political Party, Politician, Public & Government Service
Community	Community Organization, Community, Community Center, Community Service
Advocacy group	Interest, Labor Union, Environmental Conservation Organization
NA	no categories extracted

Table 4.5: Ads underreported by advertisers or volunteers that are not about political figures, elections or social issues (according to workers).

Advertiser	Text	Workers' label	*fr*	disc.
Ads underreported by advertisers: strong political ads and political ads w/o. disclosure				
Citizens Against Lawsuit Abuse	Frivolous lawsuits are clogging our courts. Want to help tell trial lawyers enough is enough? Join Citizens Against Lawsuit Abuse (CALA) today!	non-pol	1	w/o.
The Young Turks	Support independent investigative journalism while looking fly AF!	non-pol	1	w/o.
Mikey Weinstein, MRFF	MRFF Op-ed: Anti-Theist Airman Memorialized by Air Force Unit with Image of Jesus	non-pol	0.75	w/o.
Voices for Refugees	Torrential monsoon weather has hit Rohingya refugee camps in Bangladesh, destroying 273 family shelters already. Every donation helps us to reach those most vulnerable with emergency support and help to rebuild, reinforce and secure their shelters.	non-pol	0.6	w/o.
Ads underreported by volunteers: non-political ads and marginally political ads w. disclosure				
Grist.org	Lettuce introduce you to the future of your arugula.	non-pol	0	w.
Heifer International	Truth bee told, not everyone can get these 7 questions right. Test your bee smarts and unlock a 50 cent donation for Heifer	non-pol	0	w.
EveryLibrary	Join hundreds of thousands of Americans who love libraries!	non-pol	0	w.
Mercy For Animals	Animals at factory farms suffer in unimaginable ways. They are cruelly confined, abused, neglected, and mutilated. Please support our work to stop this torment.	non-pol	0	w.
The Christian Science Monitor	He need to address corruption in the Arab world is urgent. But if new initiatives are simply politically expedient – as many citizens suspect – they risk only fueling distrust and suspicion.	non-pol	0	w.

Table 4.6: Civil and social rights ads in different ad groups.

Advertiser	Text	*fr*	disc.
	Strong Political		
AFSCME 3299	Stand up for immigrant families. Tell UC to cancel its contracts with ICE collaborators now!	1	w.
SEIU	Too many of us are still paid less for the same work. That's why we need a union.	1	w.
Fight Back	We're fighting for better healthcare and equal pay.	1	w/o
ACLU	Just a few weeks ago, the Supreme Court ruled that the First Amendment forbids religious hostility by the government. If only it applied that standard to the president and his Muslim ban.	1	w/o.
	Political		
CREDO Mobile	"We can transcend the darkness of this moment by joining the struggles of past and future freedom fighters. That is how, when we reach the end of our lives and look back on these heady moments, we will find peace in the knowledge that we did our best." – Ady Barkan.	0.82	w.
Granite State Progress Education Fund	Stop anti-abortion shame, stigma and hate from New Hampshire politicians. Sign the petition to support abortion access for all Granite Staters!	0.83	w.
Physicians for Human Rights	Doctors and nurses are standing up against human rights abuses across the world. Join our community and learn more about our work	0.75	w/o.
International Rescue Committee	Women and girls in crisis zones face discrimination, violence, and a lack of equal opportunities. Learn how we're working to change that.	0.66	w/o.
	Marginally Political		
Boston Rescue Mission	It's tragic to be all alone and hungry. Your gift can bring hearty, nutritious meals to men and women who struggle with homelessness.	0.2	w.
No Kid Hungry	Giving Tuesday is coming, and you can help end childhood hunger in America. Our partner, Citi, will match all donations up to $100,000!	0.3	w.
International Rescue Committee	Yemen is facing the largest humanitarian crisis of our time: millions of children are at risk of starvation and a deadly cholera epidemic remains a serious threat. And it's about to get worse if we don't step up our efforts now.	0.2	w/o.
World Food Programme	Authorities in Yemen are blocking aid. Millions are suffering the consequences. Add your name today to keep aid flowing into Yemen	0.33	w/o.
	Non-political		
Save the Children US	There's still time to give during the 48 Hours of Giving! Your gift in support of the Center for Girls will be matched 2x by an anonymous donor – but the match ends at midnight Saturday	0	w.
United States Holocaust Memorial Museum	It's more important than ever that people understand the dangers of unchecked hatred. In this time of growing antisemitism at home and abroad, we all have a responsibility to keep the history of the Holocaust alive. Can we count on you?	0	w.
Covenant House International	TRIPLE your impact on precious young lives. Give now to help ensure that Covenant House keeps its pledge to welcome ALL homeless youth who come through our doors and love them unconditionally	0	w.
Nashville Rescue Mission	Water can be life-saving when summer's heat is at its worst and there's no escape. Helping is easy—and it won't cost you a thing	0	w.

Advertiser	Text	*fr*	disc.
Strong Political			
National Parks Conservation Association	No organization has won more victories for the national parks over the past century than NPCA - but we can't do it without you. Please donate to protect our nation's magnificent public lands.	1	w.
Conservation Northwest	ACTION ALERT: Yesterday, the U.S. House of Representatives passed a budget bill that would block funding for grizzly bear restoration in the North Cascades. Use the links below to send your elected representatives a quick message to ensure Congress provides the funding bears need!	1	w.
Ocean Conservancy	Offshore oil spills can harm marine life, devastate ocean environments and risk the livelihoods of coastal communities.	1	w/o.
Coloradans for Responsible Energy Development	Colorado's first-in-the-nation oil and gas regulations work to protect our communities.	1	w/o.
Political			
Care2	Botswana is considering lifting the ban on hunting elephants. We must act NOW and convince Botswana to maintain their stance on protecting these endangered elephants from poachers!	0.55	w.
American Bird Conservancy	The Endangered Species Act is under attack. Despite the fact that 99% of species shielded by the Act — including Bald Eagles and California Condors — have avoided extinction, opponents in Congress are threatening to undermine this bedrock environmental law. Add your name to ABC's petition and tell the government to help protect endangered birds now	0.83	w.
NowThis	Women are equally impacted by climate change, and it's critical that we have them equally involved in the solution	0.66	w/o.
NRDC	Plastics never break down. And that's becoming a real problem for those of us that depend on the Gulf of Mexico and Mississippi River	0.8	w/o.
Marginally Political			
Defenders of Wildlife	The support from our donors has helped us win many battles for wildlife, but there is always more to be done. Our love of animals is endless, so we are ready to fight tirelessly for imperiled wildlife that can't speak for themselves. Support Defenders today and help us continue the fight for wildlife!	0.33	w.
National Audubon Society Action Fund	Climate change threatens the birds we love. Sign up and we'll alert you to actions you can take to protect birds and the places we all need.	0.17	w.
Potomac Conservancy	Trees are nature's Brita filters! For just $33, we'll plant a native tree along the Potomac River to help filter out water pollution. Plant a tree today!	0.36	w/o.
Climate Reality	Last year, 39 million people tuned in to 24 Hours of Reality to learn what climate change is doing to our planet and how we can solve it with the solutions in our hands today. Help us make 2018's show even bigger!	0.33	w/o.
Non-political			
National Park Foundation	Working together, you can help us have a powerful impact on our spectacular national parks. Your support right now will go to work immediately to protect the places that matter most for future generations.	0	w.
The Nature Conservancy	The challenges facing our natural world are growing every day. Please, make a tax-deductible gift to give nature and wildlife a future.	0	w.
Oceana	Sea lions are drowning in mile-long "walls of death" off the California coast. Let them die... or help us save them	0	w.
IFAW	IFAW protects animals and the places they call home. With your help, we can continue to make a difference. Let's get to work.	0	w.

Table 4.7: Environmental politics ads in different ad groups.

Table 4.8: The average accuracy and true positive rate for a 1% false positive rate for the four models. Each classifier is evaluated over its specific training-validation dataset.

Classifier	Accuracy	TPR for 1% FPR
M_{op} model	96% (+/-1%)	92% (+/-5%)
M_{sp} model	97% (+/-1%)	96% (+/-2%)
M_{mp} model	95% (+/-2%)	90% (+/-4%)
M_{1p} model	95% (+/- 3%)	85% (+/-8%)

Table 4.9: The fraction of ads labeled as political by the four models in different groups of ads in S_2.

	# ads	M_{op}	M_{sp}	M_{mp}	M_{1p}
AdAnalyst(non-political)	1000	1.3%	1.1%	1.6%	1%
official political ads	25k	95%	90%	97%	91%
strong political ads	13k	97%	97%	98%	94%
political ads	8.7k	94%	91%	97%	90%
marginally political ads	3.7k	86%	67%	89%	84%
non-political ads	0.6K	85%	60%	89%	83%

Table 4.10: p_{value} from the Mann-Whitney U test for different groups of ads to verify NH_1 (Exp_1: no def. vs Exp_2: def) and NH_2 (Exp_1: no def. vs Exp_3: fraction).

	NH_1	NH_2
strong political ads	0.4	0.46
political ads	0.13	0.29
marginally political ads	**0.0003**	0.38
non-political ads	0.06	0.5

Chapter 5

Detection of policy-related political ads

This chapter covers the work in the paper: WWW'23 [85].
This work was done in collaboration with my supervisor Oana Goga, Romaissa Kessi (intern supervised by Oana Goga and myself) and Maximin Coavoux (CNRS).

In this chapter, our focus is on political ads that are directly associated with policy matters. Understanding which policies politicians or organizations promote and to whom is essential in determining dishonest representations. We proposes automated methods based on pre-trained models to classify ads in 14 main policy groups identified by the Comparative Agenda Project (CAP). We discuss several inherent challenges that arise. Finally, we analyze policy-related ads featured on Meta platforms during the 2022 French presidential elections period.

5.1 Data collections

5.1.1 Dataset of political ads

We collected political ads featured on Meta's core advertising platforms during the 2022 French presidential election period (Jan 1st, 2022, to June 15, 2022). To do so, we built a data collection pipeline that, each day, retrieved the Meta's Ad Library daily report [62]. This report contains information about advertisers (id and page name) who published ads, the number of ads, and the money spent. We then used the advertisers' ids to retrieve all ads about social issues, elections, or politics using the Ad Library API.

For each ad, Meta's ad library provides the: For each ad, Meta's ad library provides the:

- *creation time* of an ad.
- *creative body*–a text of an ad.
- *bylines*–the information about who paid for an ad, that advertisers are required to provide.
- *demographic distribution* – information about the age and gender of people reached by an ad.
- *region distribution*–distribution of people reached by an ad over regions in France.
- *impressions*–a field that shows the number of times the ad created an impression.
- *language*–the list of languages of the texts of the ad.
- *currency*, that was used to pay for an ad.
- *spend* – the amount of money spent running the ad as specified in currency.

Meta does not provide an exact number for impressions and spend, only an estimated range. For future analysis, we averaged these ranges.

In total, we collected 91 865 unique political ads across 9 063 pages. We filtered only ads in French which lead to 76 886 ads. Since the Ad Library does not provide exact values of expenditures and impressions but intervals of values, we averaged these ranges and estimated the number of impressions to be around 4 billion (3 799 324 537) and 20 million euros spent (20 679 225).

5.1.2 Codebook for policy categories

One of the most important decisions is how we want to label our data since no law defines social issues in each country which makes the task a bit confusing. The European Union's list established by Meta Ad Library contains 8 social issues which makes it limited and does not cover all the possible topics that can be addressed in political ads. Our goal is to define a set of complete and relevant social issues to be taken into account in the rest of the study.

On the theoretical side, we needed to know what level of granularity is required for this task? Is it enough to understand that the ad is about Human rights? Or do we want to understand precisely what right it is? As we increase the granularity of

the taxonomy, our need for data increases for the algorithm to train correctly on each category.

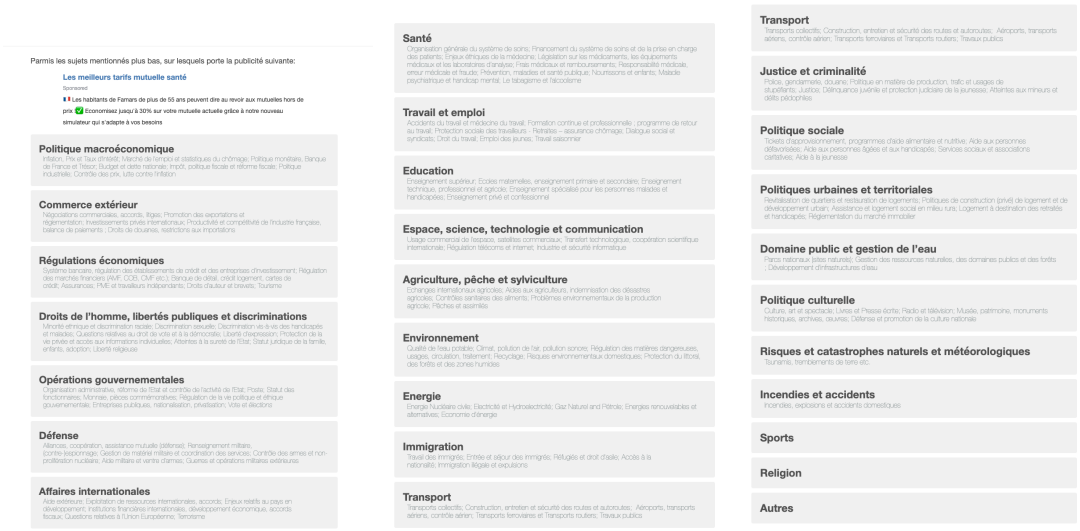
Although these tasks arise in many research applications in the political area, remarkable progress has yet to be made. The literature in political sciences offers two noteworthy efforts for analyzing written political text: the Comparative Agendas Project (CAP) and the Comparative Manifesto Project (CMP). They are large-scale data collection efforts that gather and code information about the political processes of governments around the world based on the content of the texts. The effort is made by research groups in multiple countries from various disciplines and across multiple decades. These efforts have been allowing researchers, students, policymakers, and the media to study political trends over time and across countries.

CMP's main goal is to archive and analyze the content of the electoral platforms of democratic countries from the end of the Second World War. For this CMP has proposed and is currently maintaining and updating a taxonomy that currently contains 54 categories [57]. The CMP codebook aims at capturing political parties' ideological positions on a left-right scale, hence, focusing on ideological goals. The CMP data collection classifies political parity manifestos across multiple countries. For France, the CMP dataset contains 7 977 units of labeled text (long documents are split in text units that are labeled independently).

CAP was created with the idea of tracking the attention of the government to particular policies. CAP creates a classification system that brings together a large number of political activities (e.g., bills, parliamentary debates, journalistic accounts) under a single theme with a taxonomy that counts 28 major topics (tab. 5.1) and 250 subtopics (e.g., waste is a sub-topic of the environment) [12]. The CAP codebook aims at capturing policy attention, and hence it aims at being comprehensive in the policy categories they propose. Contrary to CMP, CAP does not consider left-rights parties' positions and ideologies.

We decided to work with the CAP taxonomy. First, CAP's coding scheme focuses on the policy content and instruments, not political ideology, which we believe is more informative to study policy attention across demographic groups and candidates. Secondly, CAP's coding scheme aims to comprehensively cover topics of interest across countries (e.g., it does not miss important policy issues that might not exist in the U.S. but are essential in Vietnam). In contrast, the CMP codebook does not aim to be comprehensive. Finally, the CAP dataset has much richer data sources than CMP, which is only based on party manifestos.

Figure 5.1: Screenshots of an ad annotation task.



CAP dataset For France, the CAP dataset contains 36 658 units of labeled text. The dataset contains text units from sources such as laws, government communications, decrees and bills, sentences from all major party manifestos for general legislative elections in France. Even if the labeled data is collected in a different domain than ours, we use the CAP DATASET for training our classifiers. This type of training is called cross-domain transfer from a related domain and has been shown to work on other domains [68].

5.1.3 Data labelling procedure

To obtain labeled data, we hired human annotators to manually annotate political ads according to the 26 main CAP policy categories (tab. 5.1). To account for political ads that are not policy-related we add an "Other" category.

We encoded the survey using Qualtrics. Each survey contains one information page, followed by 1 page of *task understanding tests*, then followed by 20 pages of texts of the ads to be labeled. Each ad page contained an ad's advertiser and text, followed by a list of 26 policy categories to choose from (fig. 5.1). Going through a list of 26 policy categories is a hard task for workers. We pre-tested the survey with colleagues and workers to make the task more digestible. The survey version with the policy categories in bold, and short descriptions underneath was the most clear. We gave Qualtrics a list of 5 000 texts of the ads, and we instructed it to randomly pick 20 texts to populate the survey at each instantiation. We then launched a study on Prolific where we redirected workers to the Qualtrics survey. The only requirement

for workers was to be fluent in French. The survey took an average of 17 minutes to be completed. To determine the price to pay the workers, we took a reference payment of 7 pounds per hour (as suggested by Prolific). In total, we had 762 annotators.

We made sure that at least three different people annotated most ads to ensure the reliability of the assigned labels. Ads, that did not get three labels due to uneven Qualtrics' randomization mechanism, were deleted from the data set. We discarded all the answers from workers that took less than 4 minutes to complete the survey. As a result, the final set of labeled data consists of 4 465 ads. We selected the first three votes for ads with more than three annotations.

Using these labels we created two labeled datasets:

VM dataset This dataset considers the voting majority. For each ad, we only keep the policy categories selected by two or three annotators. In case annotators agree on more than one policy category, we keep all of them. There are 3 784 political ads (out of 4 465) for which at least two annotators agreed on at least one policy category. We discard the ads for which there is no agreement from the dataset. 30% of the ads are labeled with more than one policy category. Table 5.1 shows the number of ads per policy category in VM DATASET. We selected 5 000 ads randomly. As a result the imbalanced distribution reflects the attention different policy categories were given during the French presidential period

We represent this dataset following the one-hot encoding, i.e. our data are in the form of a matrix M with :

$$M_{ij} = \begin{cases} 1 & \text{if >= 2 annotators chose theme j for ad i} \\ 0 & \text{else} \end{cases} \tag{5.1}$$

For the test dataset, to deal with the imbalance in the policy categories and ensure that we test on a reasonable proportion of each class, we randomly took from the VM DATASET 100 ad texts per category to form the test set. We ignored the categories for which we have less than 90 ad texts. The test dataset contains 736 ads and we will call it the VM TEST DATASET in the rest of the chapter. We divided the rest of the data into training (2 160 ads) and validation (241 ads). This training data does not capture disagreement between annotators.

DISTRIB dataset To take into account all annotations, we create a second dataset that contains the distribution of annotations on policy categories. Prior research [37] has shown the empirical benefit of predicting soft labels, i.e. probability distributions on annotators' labels, as an auxiliary task to take into account annotators'

Table 5.1: Number of labeled ads on Prolific per policy category.

Policy category	Number of ads
Environment	683
Human rights	623
Cultural policy	469
Others	403
Health	374
Social policy	340
Energy	318
Government operations	311
International affairs	258
Work and employment	189
Macroeconomic policy	185
Education	146
Justice and criminality	136
Economic regulations	132
Urban and territorial policies	115
Immigration	96
Transport	69
Agriculture	69
Technology and communication	64
Defense	54
Religion	52
Foreign trade	40
Sports	38
Risk and natural disasters	22
Fires and accidents	3
Public domain and water management	0
Local and regional policy	0
Obituary	0

disagreement. The DISTRIB DATASET contains all the 4 465 previously annotated advertisements but considers soft labels. The matrix representation is done as follows:

$$M_{ij} = \begin{cases} 0.3 & \text{when 1 annotator selected category } j \text{ for ad } i, \\ 0.6 & \text{when 2 annotators selected category } j \text{ for ad } i, \\ 1 & \text{when 3 annotators selected category } j \text{ for ad } i. \end{cases} \tag{5.2}$$

We use the DISTRIB DATASET for training and validation, but not for testing. We split DISTRIB DATASET in train set (4 000 ads) and validation set (370 ads).

5.1.4 Analyzing annotation quality

While we took several steps to make the labeling task as easy as we could for workers, we still observe a lot of disagreement on the policy categories chosen by different

workers: on 16% of the ads annotators did not agree on *any* policy category. One reason for the observed disagreement could be due to the limited comprehension of the assignments by workers that try to perform tasks as fast as possible. Another reason might be the intrinsic difficulty of the task, i.e., even experienced annotators with a lot of time on their hands would disagree on the policy category [11]. To assess the quality of the Prolific annotations, two expert annotators (the Ph.D. students working on the project), annotated independently 50% of the VM TEST DATASET (431 ads). The two experts disagreed on 10% of the ads. After discussions and reading the codebooks several times, the expert annotators agreed on at least one policy category for the ads they initially disagreed on. In what follows, we refer to their annotations as *gold labels* and the corresponding dataset as GL TEST DATASET. In GL TEST DATASET we only keep the policy categories the two expert annotators agreed on.

Inter-annotator agreement measures are widely used to quantify the reliability of data annotations [5], or to establish an upper-bound on a systems' performance [2]. Table 5.2 shows the pair-wise Cohen Kappa between the final labels of Prolific workers (the VM DATASET) and the final labels of experts. There is a fair agreement (>0.3) for all categories, but a substantial agreement (>0.6) only for five categories. We observed by looking at the ads on which there is disagreement that they tend to have more labels from either experts or Prolific workers. To validate this intuition, Table 5.2 shows the inner-annotator agreement separately for ads with 1-2 categories (208) and ads with more than 2 categories (223). We see that for ads labeled with only 1 or 2 policy categories the agreement is substantially higher than for ads labeled with more than 2 categories. Hence, ads that relate to multiple policy categories are more confusing and lead to disagreement. However, we do observe substantial and almost perfect agreement on the rest of the ads.

To dig deeper into disagreement, Table 5.3 shows the classification ratio assuming that our golden labels are the real labels and the Prolific labels are predictions. The "Social policy" category has the highest number of false positives (small precision), while the "Economy" category has both high false positives (small precision) and high false negatives (small recall).

Table 5.4 shows examples of ads for categories that are false positives and false negatives. On reason for false positives is because Prolific workers interpret more loosely the 26 policy categories. For example, the ad: "*The situation on the Ukraine - Russia border is more than tense. Far be it from me to think that my opinion on this subject is particularly relevant. However, I am convinced that by turning to past history, we can try to shed light on certain points of this burning issue.*" was labeled

Table 5.2: Agreement between gold labels and Prolific labels for all ads, for ads with 2 or less policy categories and for ads with more than 2 policy categories.

Ads	all	1-2 policy cat.	>2 policy cat.
International affairs	0.62	0.68	0.47
Energy	0.75	0.88	0.33
Government operations	0.58	0.67	0.31
Cultural policy	0.68	0.81	0.22
Social policy	0.38	0.44	0.19
Health	0.68	0.8	0.41
Human rights	0.49	0.72	0.12
Environment	0.61	0.73	0.27
Economy	0.34	0.47	0.04

Table 5.3: Accuracy when the gold labels are considered ground truth, and the Prolific labels are considered predictions.

	Prec.	Rec.	F-1	Support
International affairs	0.61	0.74	0.67	50
Energy	0.77	0.81	0.79	68
Government operations	0.74	0.58	0.65	85
Cultural policy	0.84	0.65	0.73	80
Social policy	0.39	0.56	0.46	48
Health	0.68	0.77	0.72	56
Human rights	0.45	0.78	0.57	49
Environment	0.62	0.78	0.69	78
Economy	0.40	0.45	0.42	53

as being related to "Economy", "Human rights" and "International affairs" by Prolific workers but was only labeled as "International affairs" by experts. False negatives seem to happen when experts label ads with multiple categories, while Prolific workers label the ads with only a subset of categories. This might happen because Prolific workers try to limit the time they spend to label an ad and once they find a few relevant categories they go to the next ad. To check incomprehension in the task, we look at differences in the confusion matrices of Prolific workers and experts. The confusion matrix of Prolific workers' labels (fig. 5.2 shows that a higher number of ads is labeled as both "Energy" and "Environment" as well as "Social policy" and "Human rights", while the confusion matrix of gold labels (fig. 5.2 displays a lower intersection. Hence, one reason for disagreement is that the some workers do not see clearly enough the difference between "Energy" and "Environment" as well as "Social policy" and "Human rights".

Table 5.4: Examples of ads that caused disagreement between Prolific workers and experts.

Category	False positive	False negative
International affairs	Do you like Portugal? You are going to love 2022. More than 200 events to celebrate Franco-Portuguese friendship: music, cinema, visual arts, theatre, cinema, literature, gastronomy. discover contemporary Portugal! Support the France-Portugal 2022 Season and don't miss any event by subscribing to the page! **experts' label**: cultural policy; **prolific workers' label**: international affairs.	Afghans, Syrians, Sudanese... More than 26 million refugees have fled violence and persecution around the world.26 million, but as many unique stories, life paths and future projects. For 50 years, France Terre d'Asile works to defend the right to asylum and accompanies those who seek protection in France. We need you to continue! **experts' labels**: international affairs, human rights; **prolific workers' labels**: social policy, human rights.
Energy	It's official, today we say goodbye to winter and hello to spring! What if we take advantage of this new season to take care of nature? **experts' labels**: environment; **prolific workers' labels**: energy, environment.	The gas we consume today allows Putin to finance his war. Tomorrow, we will have to manage to do without it. But right now we can bring down the temperature... and the bill. For Ukraine, I'm wearing my #PatrioticSweater and turning down the heat. **experts' labels**: energy, international affairs; **prolific workers' labels**: economy.
Economy	With the crisis in Ukraine, some states wished to join the EU, in particular to prevent the conflict from being exported to their borders. Concretely, how can a country join the EU? **experts' labels**: international affairs; **prolific workers' labels**: international affairs, economy.	Banks must stop massively financing, without our knowledge, fossil fuels and polluting industries that aggravate global warming, the decline of biodiversity and therefore the living conditions of people. Sign the manifesto for a sustainable and transparent bank and regain power over your money. **experts' labels**: energy, environment, economy; **prolific workers' labels**: environment.

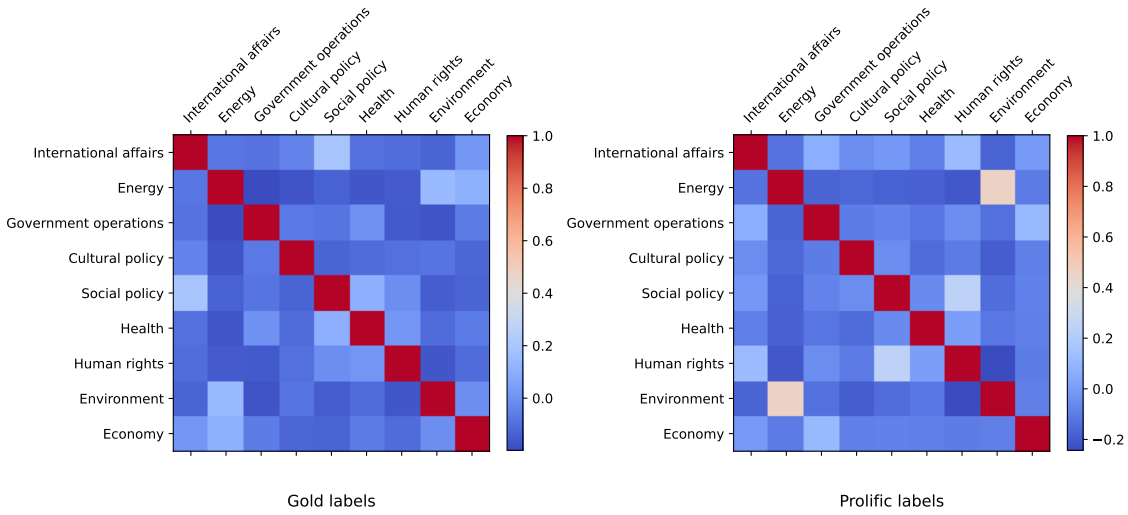


Figure 5.2: Policy category heat maps for Gold and Prolific labels.

5.2 Classification models

The task of detecting ads related to different policy categories corresponds to a *multi-label classification problem*. In multi-label classification, the training set consists of instances associated with a set of labels. In our case, we assume that an ad may refer to multiple policy categories. We built several classifiers where we tested different training datasets and hyperparameter configurations of both traditional supervised methods and recent methods based on large pre-trained language models [23]. For each classifier configuration, we build one *multi-label classifier*. We also tried One vs. All methods (i.e., having one classifier for each policy category instead of one multi-class classifier); however, it led to significantly worse results, probably because One vs. All methods do not consider any underlying correlation between the classes.

Training data.

Ideally, we would have large amounts of labeled data annotated by domain experts. Due to the unavailability of such dataset, we exploit training data that is less clean but easier to collect. Building policy detection algorithms without spending months to collect gold labels is practical in the real-world, especially if we want detection algorithms that work across languages and elections.

We instantiated three training sets based on the dataset described in Section 5.1: VM DATASET, CAP DATASET, and DISTRIB DATASET.

Data prepossessing: Prior to training, we remove links and emojis from the text of the ads. In addition, for supervised models, we also delete stop-words and punctuation signs.

Data augmentation: We use a classical data augmentation approach based on back-translation to increase the training set. Back translation consists in automatically translating an input text into another language, and then translating it back to French. The resulting text should be a paraphrase of the original text (if it is not identical to it) and can be used as a synthetic training example with the same label as the original text. We apply back translation with 40% of the ads from the train set for each category as a pivot language, and augment the training datasets to 2 542 examples (from 2 160 examples).

Supervised models. As a baseline, we chose two popular supervised models for our task: SVM [22] and Random Forests [13]. To convert the words of our data to numeric representations, we tried three vectorization techniques such as bags of words, hashing vectoriser, and TF/IDF. The last one, being the one that outperformed the others, was chosen for the rest. We used grid search to calibrate the hyperparameters of SVM and Random Forest using 10-fold cross validation.

Pre-trained language models. We use classifiers based on pre-trained language models since they are the current state of the art in text classification [23, 52]. Large language models such as BERT [23] are pre-trained on a very large amount of raw unlabeled texts (typically tens of Gb) with a self-supervised objective. They provide good-quality representations for words and sentences. Pre-trained language models have the advantage of working well with limited labeled examples thanks to the richness of the sentence representations they provide. For French, the language we are dealing with, there exist two main pre-trained models: **CamemBERT** [58] and **FlauBERT** [50]. Both models are based on the BERT architecture [23], and have been trained with a masked language modeling (MLM) objective. CamemBERT is trained on 138 Gb of textual data crawled from the web, and FlauBERT is trained with 71 Gb of data from diverse sources, including crawled data and Wikipedia.

In order to perform classification, we use the pre-trained language model to extract a vector representation for the input text, and feed this representation to a linear classification layer with a sigmoid activation. We obtain a vector $\mathbf{p} \in [0,1]^9$, such that $\mathbf{p}_l = p(l|\text{ad})$ interprets as the probability of label l given an ad: the probability of each label is modeled independently. In other words, each policy category has its own binary classifier which is siutable for our multi-label classification. Then, for a given ad, we assign it all labels l for which $p(l|\text{ad}) > t_l$ where t_l is a threshold. As a result, a single text can be assigned any number of labels (0, 1, 2 or more). A typical threshold is to use 0.5 for every category. The threshold value can also be used to control the trade off between higher recall and higher precision (sec. 5.3.2).

Training. We optimize a binary cross-entropy loss function to train the model, for 4 epochs and a learning rate of 2e-5. For better convergence, we used a linear-decreasing learning rate during optimisation and a batch size of 8. Our implementation uses the `transformers` library [45] for FlauBERT and CamemBERT pre-trained models.

5.3 Model evaluation

In this section, we evaluate the models described in sec. 5.2 in order to select the best model for the next part of our study (sec. 5.4) and we provide an analysis of their behaviour to better understand the limitations. In particular, we assess the effects of the classification algorithm, and the training dataset.

Evaluation sets Due to the category imbalance in the data (sec. 5.1), some policy categories are very infrequent. Therefore, in order to make evaluation more reliable, we build and test classifiers on a subset of the data with the 9 policy categories with more than 200 labeled ads, namely: *environment*, *international affairs*, *energy*, *civil rights*, *government operations*, *health*, *social policy*, *cultural policy*, and *economy* (which includes foreign trade, macroeconomic policy and economic regulations). We call this evaluation dataset VM-9 TEST DATASET. It contains 736 ads whose labels are obtained by a majority vote from Prolific workers. In addition, we evaluate models on GL TEST DATASET, the subset of VM-9 TEST DATASET for which we have labels provided by domain experts (sec. 5.1). GL TEST DATASET contains 431 ads.

After performing model selection on VM-9 TEST DATASET and GL TEST DATASET, we retrain the best classifier over a training set with 14 policy categories by adding categories that have more (or close to) 100 labeled ads. We will base our study in sec. 5.4 on this retrained model.

Evaluation metrics For each of our experiments, we report traditional evaluation metrics for text classification, namely: precision, recall and F_1 score for each category, as well as a micro-average across the whole test set of these metrics.

5.3.1 Results

Comparing classification algorithms We first train the four models we used (i.e., SVM, Random Forest, FlauBERT and CamemBERT) on VM DATASET and we report the accuracy of the best configuration of each classifier, as selected by

Table 5.5: Accuracy across models over VM-9 TEST DATASET. The tables presents the micro-averages of precision, recall and F1 scores.

	Precision	Recall	F-1
SVM	0.45	0.40	0.52
Random Forest	0.39	0.33	0.46
FlauBERT	**0.79**	**0.59**	**0.68**
CamemBERT	0.72	0.61	0.66

10-fold cross validation. We present the results of their predictions on VM-9 TEST DATASET in Table 5.5. As expected, FlauBERT and CamemBERT outperform SVMs and Random Forests by a large margin, and obtain F_1 scores over 0.65. This is in line with current research in NLP: the pretraining on massive amounts of unlabeled data makes language models able to adapt quickly to a downstream task, even when the size of the training set is small.

In what follows, we settle on the FlauBERT-based classifier, that slightly outperformed the CamemBERT-based classifier.

Comparing training sets Models whose results are reported in Table 5.5 were trained on VM DATASET. However, recall that we also have CAP DATASET, a dataset which contains a different type of documents (sec. 5.1) but is nevertheless much larger (25.4k labeled examples). We hypothesize that the size of this dataset may compensate for the discrepancy in terms of types of documents, and that the resulting model would generalize well on our test set, achieving cross-domain knowledge transfer.

We present the results of the models trained on CAP DATASET in Table 5.6. Unfortunately, the hypothesis turned out wrong: when trained on CAP DATASET, the FlauBERT-based classifier only achieved 0.13 F_1. This could be due to the domain discrepancy between the political ads from Meta and the documents in CAP DATASET, in terms of vocabulary distribution or length (the average length of a CAP document is 36 tokens, whereas it is an 63 tokens for an ad). Moreover, in CAP DATASET, each document has a single label, whereas in the evaluation set, an ad may have several labels, leading to a distribution shift between the train and test set which might confuse the model.

Prior work has shown that soft labels might help the classifier [70, 59, 37]. Indeed, disagreement between annotators is not only due to noise, but can also contain an

Table 5.6: Accuracy across training datasets. Comparison of FlauBERT's accuracy trained with CAP DATASET, VM DATASET (majority vote labels), and DISTRIB DATASET (soft labels).

	Training set	Precision	Recall	F-1
FlauBERT	CAP DATASET	0.14	0.11	0.13
FlauBERT	VM DATASET	0.79	0.59	0.68
FlauBERT	DISTRIB DATASET	0.79	0.60	0.68

important signal. For example, if two categories are systematically prone to disagreement, they might overlap partially in their definition. This signal can be exploited by a classifier by weighting the labels in the training data by the proportion of annotators who chose a specific label for a give example, as an indication of uncertainty for the model.

We investigate whether modelling annotator disagreement helps in our case by training the FlauBERT-based classifier on DISTRIB DATASET. The resultants are presented in Table 5.6: the training with soft labels does not improve upon training on majority-voted labels.

Results per category Table 5.7 illustrates the precision, recall and F1 score of FlauBERT across the 9 policy categories. The support is the number of texts in a specific category in the test set. We observe that some policy categories such as *environment, energy* and *cultural policy* are well detected, whereas the accuracy is much lower for ads related to *social policy*. Overall, these categories with high accuracy correspond to those with a higher agreement between annotators (tab. 5.2), and conversely *social policy* and *economy* have the lowest agreement. Indeed, a low agreement indicates both that the annotations are less reliable, and that the category is harder to detect.

5.3.2 Evaluation of the final model

The previous section showed that the overlap in content between policy categories has a negative impact on the achievable accuracy. In this section, we look at how accuracy changes when we consider more policy categories. Here, we train and test classifiers using 14 policy categories for which we have more than 100 ads. We add the following policy categories: *education, justice and crime, work and employment, urban and territorial policies*, and *immigration.* Unfortunately, we do not have enough labeled data to add the 12 other categories.

Table 5.7: Accuracy across policy categories using FlauBERTtuned over VM-9 TEST DATASET.

	Prec.	Rec.	F-1	Support
International affairs	0.81	0.60	0.69	100
Energy	0.93	0.68	0.79	100
Government operations	0.65	0.43	0.52	105
Cultural policy	0.84	0.83	0.83	109
Social policy	0.76	0.19	0.30	102
Health	0.86	0.73	0.79	102
Human rights	0.67	0.47	0.55	125
Environment	0.81	0.80	0.81	150
Economy	0.75	0.49	0.59	102
micro avg	0.79	0.59	0.68	995
macro avg	0.79	0.58	0.65	995
samples avg	0.72	0.64	0.66	995

Table 5.8 shows the results across the 14 policy categories. The table shows that even for policy categories such as *immigration* and *urban and territorial policies* for which we have less than 200 ads, the classifier is able to achieve F1 scores over 0.5. The table also shows that the accuracy of the initial 9 policy categories slightly drops. Indeed, the additional categories make the task harder. A higher number of categories also leads to higher potential overlap between categories.

Controlling the precision-recall trade off For our case study, the precision is more important than the recall–it is more important not to mislabel ads with the wrong policy category than to miss some ads that are related to a policy category. Given this preference, instead of using the same threshold for each category (i.e., 0.5), we select a different threshold for each policy category.

To get the appropriate threshold for each category we performed *threshold optimization* as a fine-tuning step. The definition of the threshold is done during the validation phase by maximizing precision over recall. We, therefore, look for thresholds that give a precision close to 85% with the highest possible recall.

Table 5.8 presents the precision and recall of FlauBERT with 14 policy categories using different thresholds. Note that the precision is not always close to 0.85 since the thresholds have been defined on validation data and not on test data. In the next section we use this model for label prediction. Table 5.9 presents the precision and recall of FlauBERT with 14 policy categories using the same thresholds. In the next

	Prec.	Rec.	F-1	Support
International affairs	0.93	0.26	0.41	102
Energy	0.96	0.45	0.61	100
Immigration	0.57	0.27	0.36	30
Law and crime	0.75	0.26	0.38	35
Government operations	0.73	0.21	0.33	105
Cultural policy	0.78	0.75	0.76	110
Social policy	0.76	0.12	0.21	104
Urban and territorial policies	0.89	0.26	0.40	31
Health	0.84	0.60	0.70	101
Labor and employment	1	0.17	0.29	36
Human rights	0.79	0.20	0.32	134
Education	0.75	0.29	0.42	31
Environment	0.79	0.71	0.75	150
Economy	0.89	0.16	0.26	103
micro avg	0.81	0.38	0.51	1172
macro avg	0.82	0.34	0.44	1172
samples avg	0.52	0.43	0.46	1172

Table 5.8: Accuracy across policy categories using FlauBERTtuned over the VM-14 TEST DATASET. **Different thresholds per category.**

section, we are using the combination of these two models for analysis.

5.4 Case Study: Policy attention in the 2022 French election Ads

Political scientists and analysts have long been interested in policy attention dynamics across countries and elections [66]. However, most studies have analyzed policy attention through manual annotations of various sources of texts such as political parties manifestos, mass media, and senate hearings. As a case study, to show the practical usefulness of the classifier we developed in the previous section, we analyze how policy attention varied across candidates and different demographic groups during the 2022 French Presidential election (held in two rounds: 10 April and 24 April). We applied the FlauBERT model for the analysis with different thresholds on the 76 067 political ads we collected from Meta's Ad Library that ran between 1 January 2022 and 15 June 2022. FlauBERT model with different thresholds ensures high precision, hence, being confident that all the ads labeled about a specific policy are correct. However, the recall varies across policy categories from 0.16 to 0.75. Hence, we cannot detect *all* ads corresponding to every policy category. For this section, this is not problematic as our analysis compares policy attention in different demographic groups and across presidential candidates, and a low recall should count equally in

Table 5.9: Accuracy across policy categories using FlauBERTtuned over the VM-14 TEST DATASET. **Same thresholds per category.**

	Prec.	Rec.	F-1	Support
International affairs	0.68	0.49	0.57	102
Energy	0.92	0.59	0.72	100
Immigration	0.71	0.50	0.59	30
Law and crime	0.83	0.29	0.43	35
Government operations	0.64	0.41	0.50	105
Cultural policy	0.83	0.71	0.76	110
Social policy	0.85	0.11	0.19	104
Urban and territorial policies	0.61	0.45	0.52	31
Health	0.85	0.67	0.75	101
Labor and employment	0.59	0.61	0.60	36
Human rights	0.63	0.51	0.57	134
Education	0.88	0.23	0.36	31
Environment	0.76	0.79	0.78	150
Economy	0.89	0.23	0.37	103
micro avg	0.75	0.50	0.60	1172
macro avg	0.76	0.47	0.55	1172
samples avg	0.64	0.56	0.58	1172

all groups.[1] Out of the 76 067 political ads, our model predicted at least one policy category for 59 718 ads. Moreover, for 6 531 ads the model predicted more than one policy category and ads had in median 1 policy category.

5.4.1 Policy attention and presidential candidates

We analyze both policy attention in ads coming from the official accounts of presidential candidates and their corresponding political parties and ads that do not necessarily come from official accounts but mention a candidate's name. Remember that on Meta, anyone can be an advertiser and send political ads if they prove they reside in the same country where the ads are targeted.

There were 12 candidates in the election, and we manually found all official corresponding accounts. In France, the law prohibits, in the six months preceding an election, the use for electoral propaganda purposes of any commercial advertising process by the press or by any means of audiovisual communication [51]. Despite the law, we observed 321 ads (corresponding to 23 443 021 million ad impressions) coming from several official accounts of presidential candidates posted from 1 Jan to 24 Apr. We saw Emmanuel Macron's party "En Marche" circumventing this prohibition by

[1]Different recalls for different policy categories will be problematic if the goal of the analysis would be to determine the policy categories that most attention.

financing a few days before the elections "register to vote" ads on Facebook targeting users ages 18-24 posted on the page "La France aux urnes".[2] In addition, Eric Zemmour and Marine Le Pen (two prominent right-wing extremists) also sponsored political ads encouraging users to join their party or support them through donations.

Secondly, we identified 1598 ads that mention one of the top three presidential candidates according to votes in round 1: 1 050 mention Emmanuel Macron, 406 mention Marine Le Pen, and 142 mention Jean-Luc Mélenchon. Table 5.10 shows the policy attention of ads corresponding to different presidential candidates. To measure policy attention, we collected information from Meta's Ad Library about the number of ad impressions (i.e., the number of users that saw the ad) of each ad in our dataset (sec. 5.1). Hence, we summed up the ad impression information for all ads mentioning a particular candidate and labeled with a particular policy category. The table shows that many ads mentioning the candidates address "Government operations" which makes sense since this category describes everything related to the elections and the country's state. The distribution of ad impressions across the other policy categories is uneven across candidates. The majority of the ads that mention Macron discuss "International affairs". This can be justified by the strong involvement of the French president in the war between Ukraine and Russia. In contrast, most ads mentioning Le Pen discuss "Health" and most ads mentioning Mélenchon (besides "Government operations") discuss "Economy". Understanding how candidates represent themselves and on which policies they focus their attention, and how the large public talks about the candidates is important for mandate accountability and understanding how democracies evolve.

5.4.2 Policy attention across demographic groups

Meta's Ad Library provides information on the demographic distribution of people reached by every political ad. In Table 5.12, we use this information to study policy attention across demographic groups by investigating what are the demographic groups that are over/under targeted by the different policy categories. Each cell represents the proportion of ad impressions of ads related to a particular policy categories that have been viewed by a particular demographic group. The first line of the table (i.e. Population) represents the demographic distribution of all ad impressions in French that have at least one predicted policy. We use this as a baseline to identify over-exposure (in red) and under-exposure (in blue). A few interesting observations we see in the table:

[2]https://www.facebook.com/la.france.aux.urnes.2022

Table 5.10: Distribution of ads impressions by policy category in ads mentioning different presidential candidates.

	E. Macron	M. Le Pen	J-L. Mélenchon
International affairs	29.14%	7.37%	1.77%
Energy	0.48%	0.00%	0.00%
Immigration	0.26%	0.11%	0.00%
Law and crime	0.34%	0.00%	0.00%
Gouvernement operations	30.69%	16.00%	63.02%
Cultural policy	3.27%	0.06%	0.00%
Social policy	0.72%	0.07%	1.47%
Urban and territorial policy	0.30%	0.09%	0.04%
Health	2.75%	49.03%	1.51%
Work and employment	0.73%	0.07%	0.00%
Human rights	4.39%	0.31%	1.26%
Education	0.95%	0.37%	0.00%
Environment	10.52%	14.84%	0.00%
Economy	15.46%	11.68%	30.91%

(1) Women are under-exposed (compare to men) to ads talking about "Energy" and "Economy" and they are over-exposed to ads about "Immigration", "Social policy", and "Health".

(2) Users aged 18-24 are under-targeted to ads about "International affairs", while users over 65 are over-targeted.

(3) Users aged 18-34 are under-targeted to ads about "Immigration", while users over 45 are over-targeted.

(4) Users aged 13-24 are severely over-targeted with ads about "Law and crime".

(5) "Cultural policy", "Social policy", "Economy", and "'Human rights" are pretty evenly distributed across demographic groups.

(6) Users aged over 55 are over-targeted with ads about "Health".

(7) Users aged 18-24 are over-targeted with ads about "Work and employment".

Overall, we do see large variations in policy attention across demographic groups. This kind of imbalance may not be beneficial as it could reinforce gender and age stereotypes, and may deprive certain users from relevant information that might be important in their voting deliberation. Who received an ad depends on both the advertiser that can specify to which gender and age groups they want to send their ad, but also the ad optimization algorithms employed by Meta [1] that optimize ad deliver. To better understand who is responsible for the imbalance in policy attention,

Table 5.11: Distribution of ad impressions across regions and policy categories. * represents the region distribution of impressions for all ads in French that have at least one predicted policy.

	Regions												
	Auvergne-Rhône-Alpes	Bourgogne Franche-Comté	Bretagne	Centre-Val de Loire	Corse	Grand Est	Haut De France	Narmandie	Nouvelle-Aquitaine	Occitanie	Pays De La Loire	Provence Alpes Côte D'Azur	Ile-De-France
Population*	11.7%	4.25%	4.96%	3.78%	0.76%	8.22%	9.60%	5.12%	9.69%	10.07%	5.39%	8.89%	17.55%
International affairs	12.25%	3.76%	4.94%	3.66%	0.80%	8.32%	8.43%	4.92%	9.69%	10.07%	4.90%	9.46%	18.79%
Energy	6.31%	6.23%	1.41%	6.50%	0.10%	6.80%	15.85%	7.54%	11.73%	7.33%	6.69%	6.71%	16.81%
Immigration	12.30%	3.98%	5.09%	3.56%	0.81%	7.98%	8.23%	4.78%	9.58%	10.24%	4.79%	9.75%	18.91%
Justice and ciminalty	5.50%	1.92%	2.96%	1.74%	0.49%	3.75%	3.69%	2.51%	5.17%	5.37%	2.59%	4.74%	59.57%
Gouvernement operations	11.76%	4.03%	5.14%	3.74%	0.79%	8.47%	9.89%	5.59%	9.89%	9.43%	5.50%	8.20%	17.57%
Cultural policy	11.04%	3.60%	5.17%	3.46%	0.78%	7.23%	7.85%	4.99%	9.38%	10.98%	5.87%	9.19%	20.47%
Social Policy	11.46%	4.27%	5.05%	3.88%	0.78%	8.77%	10.53%	5.46%	9.76%	9.81%	5.22%	8.95%	16.05%
Education	10.72%	3.79%	4.86%	3.74%	0.63%	7.82%	8.90%	4.95%	9.09%	12.90%	5.25%	7.80%	19.55%
Environment	10.66%	5.39%	4.59%	4.09%	0.64%	8.33%	10.43%	5.34%	10.50%	10.06%	5.57%	8.63%	15.76%
Health	11.17%	4.09%	4.68%	3.88%	0.85%	8.63%	10.51%	5.34%	9.39%	9.46%	4.93%	9.81%	17.27%
Economy	10.55%	4.30%	9.31%	3.50%	1.29%	8.39%	10.38%	5.00%	8.66%	8.75%	4.54%	8.82%	16.51%
Human rights	13.70%	3.87%	5.29%	3.66%	0.65%	7.90%	9.42%	5.05%	9.59%	9.92%	5.51%	8.43%	17.01%
Work and employment	10.95%	4.09%	4.88%	4.41%	0.63%	8.66%	11.92%	5.61%	9.12%	8.75%	5.33%	9.59%	16.07%
Urban, territorial policy	27.08%	4.87%	2.42%	2.25%	0.46%	14.26%	6.74%	3.13%	6.75%	6.67%	2.60%	11.18%	11.60%

Table 5.12: Distribution of ad impression across demographic groups and policy categories. * represents the demographic distribution of impressions for all ads in French that have at least one predicted policy. Over-exposure (in red) and under-exposure (in blue).

	Gender		Age						
	Female	Male	13-17	18-24	25-34	35-44	45-54	55-64	65
Population* (baseline)	53.94%	46.06%	2.68%	14.24%	22.79%	18.33%	15.14%	13.12%	13.52%
International affairs	53.64%	46.36%	0.18%	6.26%	22.22%	19.83%	16.64%	15.89%	18.98%
Energy	35.61%	64.39%	0.01%	1.74%	25.76%	34.49%	27.35%	7.88%	2.77%
Immigration	65.65%	33.35%	0.18%	4.6%	16.68%	17.06%	18.88%	20.57%	22.02%
Law and crime	51.88%	48.12%	28.57%	24.52%	8.46%	10.97%	9.3%	8.33%	9.85%
Government operations	53.35%	46.65%	0.92%	30.26%	27.32	13.83%	9.7%	8.7%	9.23%
Cultural policy	51.53%	48.47%	3.17%	16.18%	23.32%	17.97%	15.63%	12.21%	11.52%
Social policy	65.34%	34.66%	0.93%	13.65%	19.74%	16.27%	14.51%	16.15%	18.74%
Education	59.19%	40.81%	11.4%	24.2%	16.6%	13.8%	9.67%	10.26%	14.79%
Environment	50.89%	49.11%	1.95%	10.78%	25.52%	21.91%	17.04%	12.17%	10.64%
Health	68.45%	31.55%	4.16%	8.62%	14.18%	18.12%	17.12%	18.42%	19.37%
Economy	44.24%	55.76%	0.01%	12.85%	22.25%	19.36%	15.67%	15.53%	14.34%
Human rights	59.31%	40.69%	8.74%	16.18%	20.09%	16.52%	13.54%	12 %	12.13%
Work and employment	61.27%	38.73%	0.84%	27.49%	19.03%	12.60%	14.79%	12.88%	12.37%
Urban and territorial policy	51.09%	48.91%	3.63%	6.32%	14.62%	15.90%	15.60%	17.65%	26.29%

it is necessary that ad platforms provide more transparency about the demographics chosen by the advertiser, and the demographics of the users the ad actually reached.

Table 5.11 shows the distribution of ad impressions across regions and policy categories. Generally, the distribution of ad impressions in different policy categories is close to the distribution of ad impressions across the whole France. Nevertheless, there are differences between regions, we see that around 60% of all ads about "Justice and criminality" were shown to the people from the Ile-De-France region; and more than 27% of the total ads about "Urban policies and territories" were shown in Auvergne-Rhône-Alpes.

5.5 Data visualisation

For data visualization, we developed a web server. This web application follows a three-tier architecture based on the RESTful (Representational State Transfer) style. It includes a presentation layer, which is the front-end interface accessible through a web browser, a business layer that serves as the API back-end server handling user requests from the front-end, and a data access layer that stores the application's data in a DBMS (Database Management System).

The system has been developed using a combination of key technologies, including:

- MySQL: A renowned relational database management system (RDBMS).
- JavaScript: A widely-used scripting programming language that plays a crucial role in creating interactive web pages and is fundamental to web applications.
- Node.js: A free software platform that utilizes JavaScript and is specifically designed for high-performance, scalable network applications.
- Express: A popular framework for building web applications based on Node.js, considered as the standard framework for server development in Node.js.
- React: A free JavaScript library developed by Meta since 2013, aimed at simplifying the creation of single-page web applications through the use of state-dependent components that generate HTML pages (or portions) upon state changes.
- Chart.js: A free open-source JavaScript library used for data visualization.

The server presents general information about all political ads sent during the election period on the home page (fig. 5.3). This includes data on the evolution of sponsored political content on a monthly basis, ad spending, and geographic breakdown.

Section **Analytics** presents a more comprehensive analysis. Subsection **General** enables users to explore detailed statistics of the advertisers, including impressions, spending, and the number of political ads sent. It also provides a breakdown of the population reached by the ads by age and gender, as well as a timeline of political ads' spending related to specific policy categories.(fig. 5.4).

First Round and **Second Round** subsections displays statistics regarding ads that mention election candidates and ads sent by election candidates during the two rounds of elections, respectively (fig. 5.5).

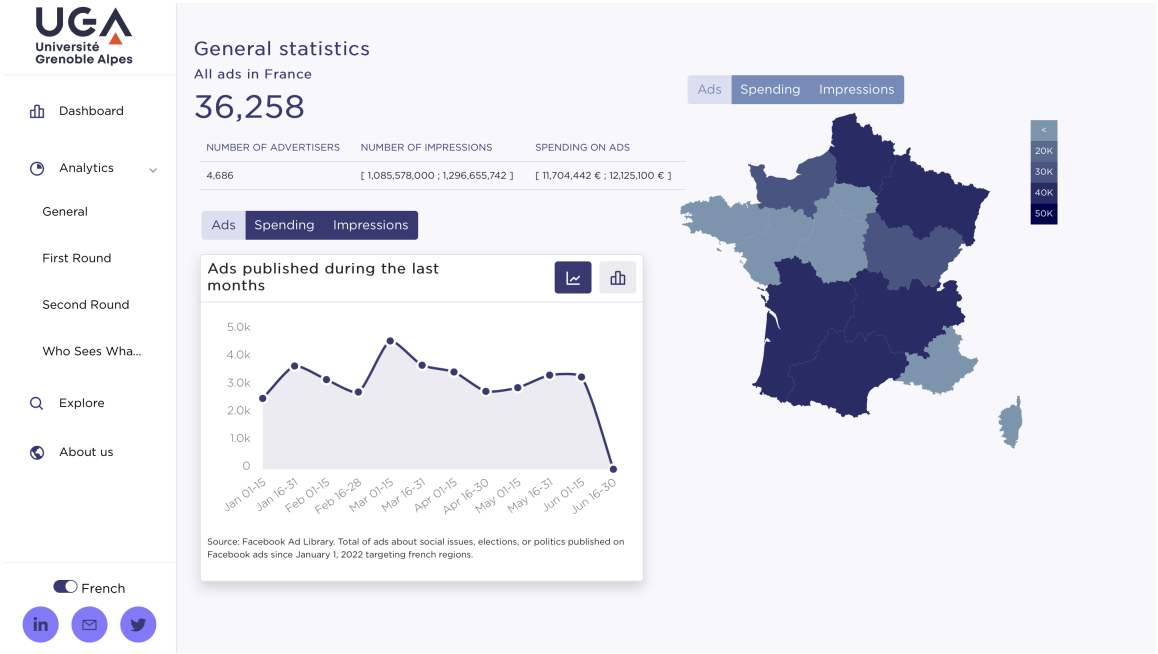


Figure 5.3: The home page of the server displays general information related to political ads during the election period.

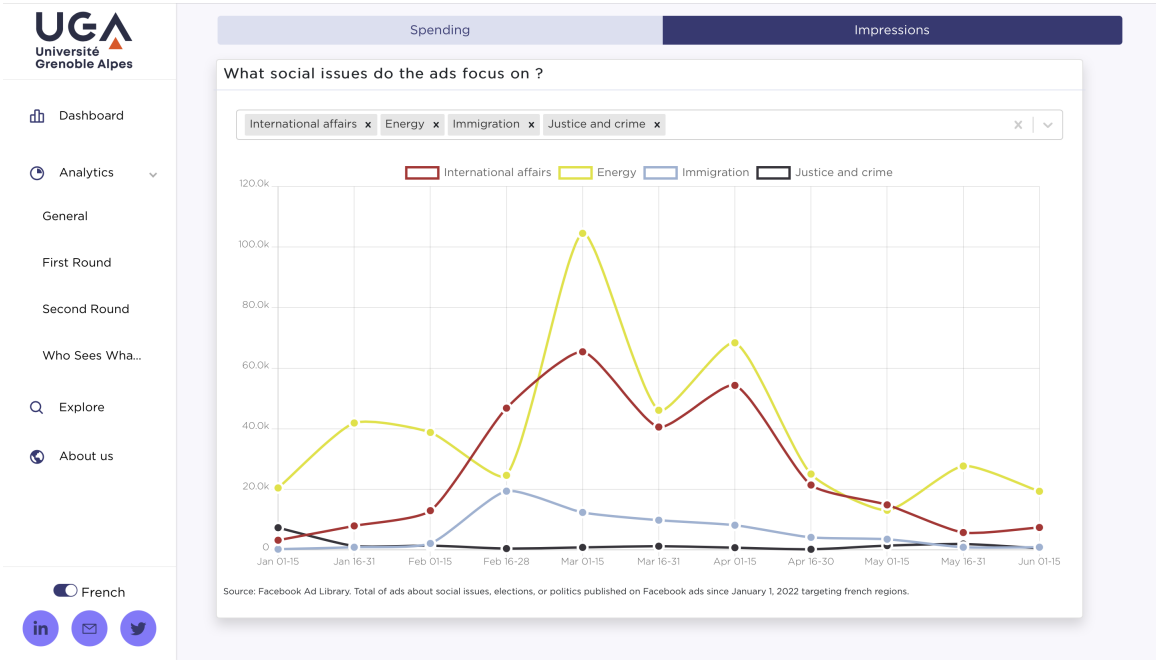


Figure 5.4: The page of the server displays the timeline of spending related to specific policy categories.

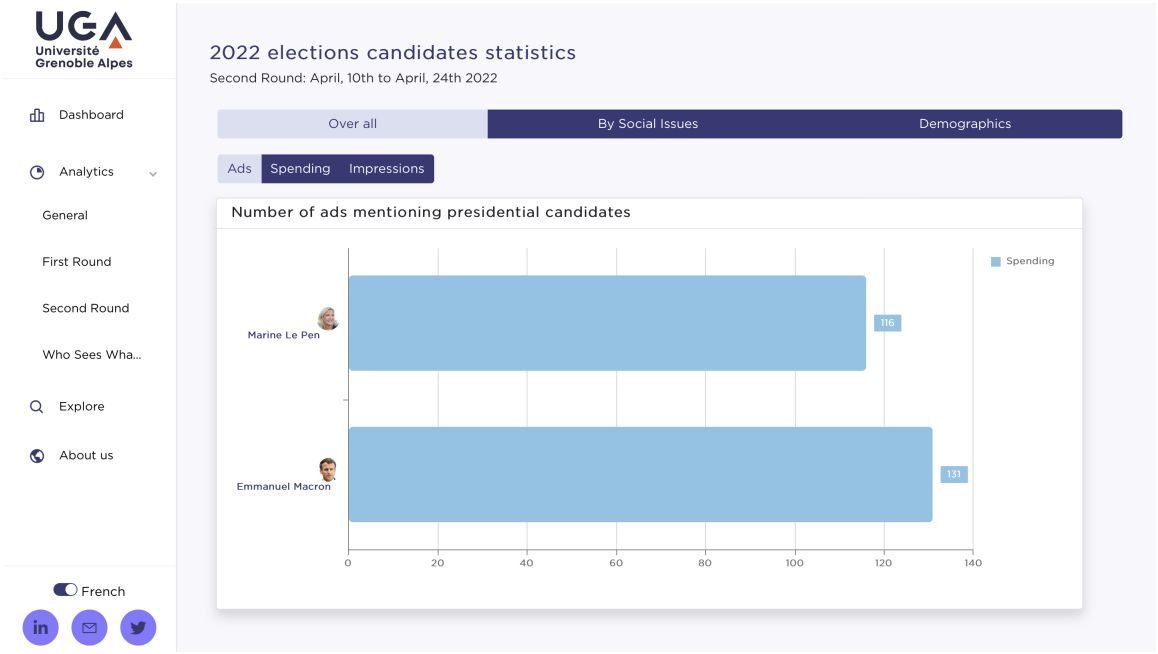


Figure 5.5: The page of the server displays spend on ads mentioning candidates during the second round of election

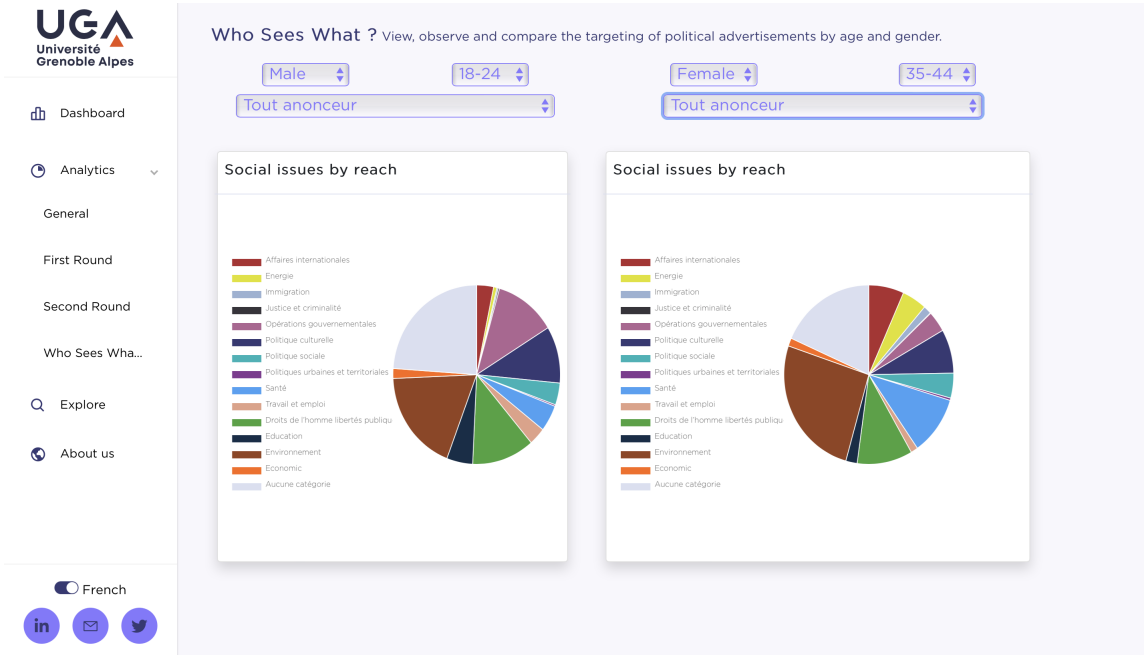


Figure 5.6: The page of the server displays the breakout of policy categories based on targeting information and advertisers.

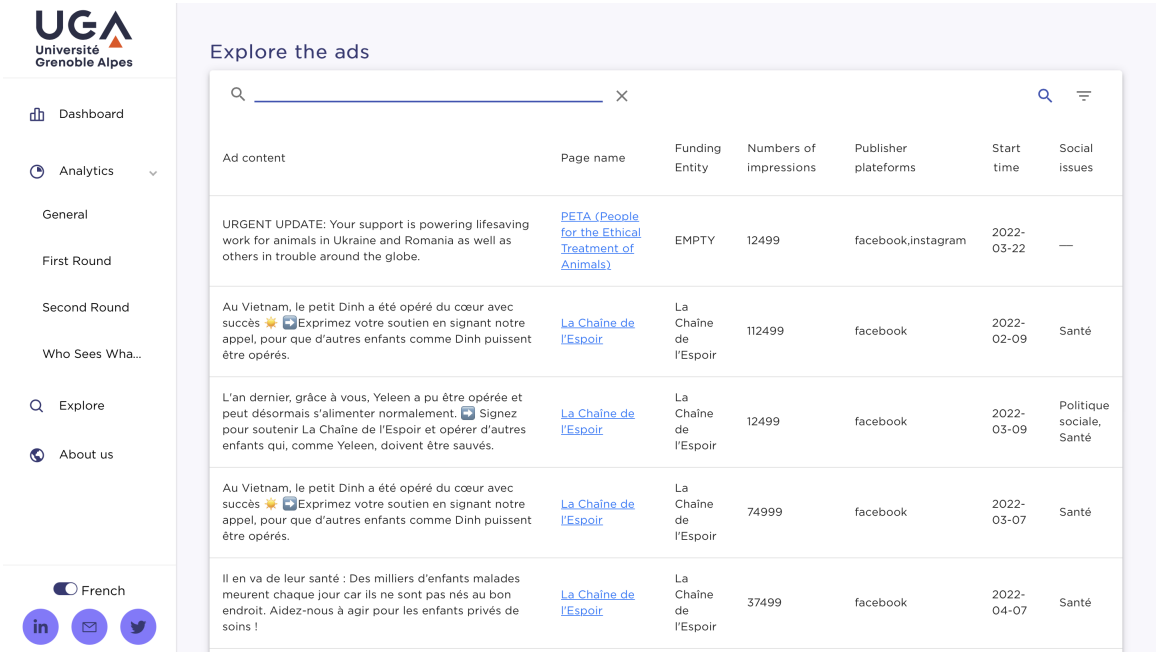


Figure 5.7: The page of the server displays filtering advertisements based on specific keywords.

Subsection **Who Sees What** presents information about the breakout of of policy categories based on targeting information and advertisers (fig. 5.6). The user can analyze and compare the content of different advertisements that target specific age groups, genders, and are broadcasted by specific advertisers, based on social policies.

Explore section is allow users to search for advertisements based on the specific keywords in the ad, and access detailed information about the advertisements, including the ad body, funding entity, and the page that posted it (fig. 5.7).

The server is available at: https://elections2022.imag.fr

5.6 Summary

New transparency initiatives from online platform and governments enable the public to access information about all ads running on these online platforms. Given the large volumes of data available, there is an increasing need for automatic methods to investigate paid political speech. This work explores automated methods to label political ads according to 14 policy categories. Understanding policy attention is important for analyzing democratic processes. Our models are able to achieve over 0.75 F1 scores for policy categories such as *environment* and *cultural policy* and F1 scores between 0.5 and 0.7 for policy categories such as *energy* and health. Overall, the

categories with high accuracy correspond to those with a higher agreement between Prolific annotators. The main culprit for disagreement are the ads who's messages relates to multiple policy categories. Confusion across certain policy categories such as "Energy" and "Environment" could be improved slightly, but the task remains complex even for experienced and trained annotators. Our methods could be used in conjunction with methods to detect sentiment and tone to identify deceiving political ads that exploit vulnerable groups of people through targeting.

Finally, we build one of the few models in the literature to analyze political content in French. Using this model we analyzed the online ads posted during the 2022 French Presidential Election. We observe significant imbalances in the policies discussed in ads that target different demographic groups. Such imbalances could affect voters deliberation and, hence, need to be taken into account when designing political ad targeting technologies.

Chapter 6

Benchmarks

This chapter covers the work in the paper: CoNEXT Student Workshop'22 [84]. This work was done in collaboration with my supervisor Oana Goga.

In this chapter, we propose a series of practical benchmarks designed to evaluate the effectiveness of political ad definitions. These benchmarks are aimed at assessing the ability of definitions to accurately identify truly problematic ads, known as true positives, which include ads with divisive messages targeting different demographic groups. Additionally, the benchmarks assess the ability of definitions to avoid capturing ads with purely humanitarian scopes, known as false positives. We evaluate two definitions obtained from online platforms and two definitions obtained from policymakers using our benchmarks.

6.1 Background

This section presents the basic information about what is the Non-Governmental Organization and main types of NGO.

The term "NGO" stands for "Non-Governmental Organization." NGOs are organizations that are not affiliated with any government and operate independently to address various social, environmental, or humanitarian issues. They are typically non-profit organizations and can be involved in activities such as advocacy, charitable work, development projects, and humanitarian aid, among others. NGOs are usually established by private individuals, groups, or organizations and work towards the betterment of society, often focusing on specific causes or areas of concern.

The World Bank classifies NGOs into two main groups [56]:

- Operational NGOs: These NGOs are involved in implementing and delivering programs and projects directly to communities or beneficiaries. They work

on the ground, providing services, and implementing development projects in areas such as health, education, agriculture, infrastructure, and social welfare. Operational NGOs are often involved in fieldwork and have a direct impact on local communities.

- Advocacy NGOs: These NGOs focus on advocating for specific causes or issues at local, national, or international levels. They work to raise awareness, promote policy changes, and advocate for the rights of marginalized groups or for environmental or social issues. Advocacy NGOs often engage in research, policy analysis, lobbying, and campaigning activities to bring about social or policy changes.

6.2 Benchmarks

This section presents four benchmarks to assess the quality of the definition of political ads and then it presents how we gather data for the benchmarks.

Agreement The first benchmark **agreement** assesses how user-friendly a definition is. Good definitions should be easy to apply to different types of ads. Moreover, almost anyone can place an advertisement on platforms like Facebook and Google after doing a few steps [33]. Hence, a definition needs to be easy to use not only for political scientists but for regular users as well. To summarise it, we want a definition that makes the data labeling process easy, and most people should agree on which ads are political and non-political after reading the definition.

Influence The main goal of any political campaign is to win an election. Political campaign staffs use different types of ads for different purposes. For instance, comparison and negative ads about an opponent are used to encourage people to change their opinion about an opposition candidate and convince them to vote for their candidate. Furthermore, ads about the candidate's background, personal beliefs, and promises are used to promote the candidate and ensure that the voters know the candidate's name. The also types of sponsored political content that encourages people to get out and vote to increase the turnout for the election. In other, advertisers can run ads with the main idea of minimizing the turnout for the elections of some active groups who will vote for the opposite candidate. All these different types of ads are trying to influence people voting behavior. It is essential to detect these ads to be sure that candidates do not break any laws with their campaigns. Access to these types of

sponsored political content can be helpful to political scientists and physiologists. To summarize everything above, the second benchmark is **the ability to capture ads that can influence people's voting behavior.**

Divisiveness To understand the meaning behind a third benchmark, it is necessary to take a closer look at the intervention in the 2016 U.S. presidential election. Russia-linked ad campaigns on Meta, where the Russian Internet Research Agency micro-targeted users with ads to interfere with the U.S. presidential election. One of the particularities of these ads is that their Click-Through Rate (ratio of users who click on a specific link to the number of total users who have seen it) is ten times higher than typical values for Meta, meaning that these ads were very effective. These ads were about polarising topics and targeted specific groups such as African-Americans or Latinos. To avoid this interference in the future, it is crucial that the definition can catch divisive ads about different social issues or election-related topics. It leads to the third benchmark: **ability to catch divisive ads across different racial, age, and gender groups of people.**

Humanitarian aid Ads about social issues are the primary source of disagreement in the definitions. In comparison, some of the sponsored political content about social problems is completely harmless for any election and have an operational focus such as promoting development projects (**operational ads**). Conversely, others have an advocacy focus as a primary task: promoting specific causes by persuading citizens (**advocacy ads**). The 'good' definition should be able **to catch advocacy ads and not label as political operational ads**.

6.2.1 Data collection and experiments

In this chapter, we conduct a evaluation of the quality of four definitions that were introduced in chapter 3. For our evaluation, we specifically chose two definitions that were proposed by government entities, such as the **European Parliament** and the **European Commission**, as well as two definitions proposed by social media platforms, such as **Meta** and **Twitter**. This selection allowed us to assess the quality of definitions from different sources and perspectives.

Meta dataset. For our first dataset, named Meta dataset, we collected all ads related to social issues, elections, or politics from the Meta Ad Library that were published in July 2022. From this collection, we randomly chose 500 unique advertisements. This

dataset serves as a representative sample of the typical posts considered as political on Meta and Instagram.

IRA dataset. The second dataset is called IRA dataset. It comprises 3,517 Facebool ads, reported to have been purchased by the Internet Research Agency, which were disclosed by the House Intelligence Committee in the form of redacted PDF files [21]. and parsed using irads [87]. Notably, these ads predominantly revolve around polarizing topics. From this dataset, we randomly extracted 100 unique posts.

NGO dataset. The last dataset, the NGO dataset, purely focuses on ads about social issues. The collection of this dataset involved a series of steps. To become an advertiser on Meta, firstly, the person needs to create a page and then select a category from the pre-defined list [29]. We use the advertiser's IDs available in the Ad library reports to collect their category using the Meta Graph API. After that, we pre-selected ads already collected based on the advertiser's category. We chose ads only published by "Charity" and "Non-Governmental organization." Finally, we randomly selected 100 unique ads from this pre-filtered set.

Each advertisement is characterized by its unique text, ensuring that there is no duplication or intersection between the datasets.

6.2.2 Experiments

To test our benchmarks in action, we analyzed the quality of four definitions based on them. We chose two definitions from social network platforms that are focused only on the content of an ad (Meta and Twitter definitions) and two from government organizations that are also focused on advertisers' intentions (European Commission and European Parliament definitions).

To do so, we set up surveys on Qualtrics where for each ad, we ask respondents questions about the ad's message [76]. We hired workers through Prolific, and we redirected them to fill out the survey [73]. Each worker answered questions about 25 random ads and each ad was labeled by three workers. We select workers who are fluent in English and live in the U.S because all ads initially were targeted to the U.S users.

We called the first part of our research the **agreement study**. In this study, we asked workers to label ads from all our datasets as political and non-political based on one of the definitions. We did a separate survey for each definition.

In the second part - the **behavioural influence study**, we asked workers these 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?

The behavioral influence study was conducted on Meta and IRA datasets.

The last study is called the **humanitarian aid 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?

The last study was conducted only on NGO dataset.

The agreement study took workers an average of 11.4 minutes to be completed, for the behavioral influence study 6.1 minutes, and for humanitarian aid study 11 minutes. To determine the price to pay the workers, we took a reference payment of 8 pounds per hour (as suggested by Prolific).

The primary purpose of the studies was to check how workers understand and apply different definitions. Unfortunately, due to the nature of the research design, we were unable to validate the responses provided by the participants. The sole criterion for excluding responses was the amount of time spent on the survey. We excluded workers who spent less than 2 minutes to complete the task.

Table 6.1 shows the general overview of the datasets, after analysing the results of our studies.

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

Table 6.1: 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%
Meta definition	62%
Twitter definition	66.7%

Table 6.2: Agreement Study. The table shows percentages of ads workers agreed on in total for all three datasets

6.3 Results

Results of the study about agreement are shown in the table 6.2. 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, Meta 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 Meta 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 Meta definition includes ads about social issues. This could be a possible reason for the Meta 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 behavioural influence study on two datasets (tables 6.3, 6.4). All four definitions performed well on the Meta dataset. All of them were able to catch more than 80% ads that could influence

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%
Meta definition	364	88.7%	83.8%	86.3%
Twitter definition	354	86.2%	80.7%	79.5%

Table 6.3: Behavioural Influence Study: Meta 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%
Meta definition	65	92.9%	88.9%	86.4%
Twitter definition	38	58.9%	54%	52.5%

Table 6.4: Behavioural Influence Study: IRA dataset

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

Table 6.5: Humanitarian Aid Study NGO dataset. Operational† means that an ad is only operational and does not have advocacy label.

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 Meta definitions did not 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 6.5. 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. Meta definition shows the most stable results on the fourth benchmark with mislabeling 9.5% of operational digital ads and catching 62% of advocacy ads.

Meta 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.

6.4 Summary

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.

Chapter 7

Conclusion and future work

Online political advertising is an excellent tool for communication between voters and candidates. Compared to broadcasting and paper advertising, its low cost allows candidates and companies with low budgets to advertise themselves to voters effectively. Not surprisingly, we have seen a significant increase in sponsored political content in recent years. It also provoked new challenges regarding the regulation and analysis of sponsored political content. In this thesis, we made three main contributions. Firstly, we examined the reliability of distinguishing political ads from non-political ones. Secondly, we analyzed policy-related ads from Meta during the 2022 French presidential election. Finally, we formulated benchmarks to evaluate the quality of the definition of online political ads. Although many scientists, including ourselves, have studied online political ads, this topic still demands further research and raises many questions, such as which ads should be regulated and how to regulate them effectively.

7.1 Summary of contributions

The thesis makes the following contributions:

Analyzing of disagreement on online political ads: We explore the reliability of distinguishing political ads from non-political ones using an empirical approach. In our analysis, we focused on three key aspects. Firstly, we analyse if people agree on what ads are political. Secondly, we investigated the characteristics of ads considered political by ordinary people. Finally, we looked at the characteristics of ads that lead to disagreement. Our study found that there is a significant disagreement among ad platforms, advertisers, and ordinary people about what constitutes a political ad,

especially regarding ads that address social issues. This makes it important to classify social issue ads as political, but it also complicates regulating political advertising.

Service to explore disagreement on online political ads: We developed a service aimed at exploring disagreement in online political advertising. Our service enables users to search for ads published or sponsored by various advertisers. Additionally, users can search for ads that received varying degrees of political votes and are related to different topics.

Model for classification policy-related political ads: We propose automated methods based on pre-trained models to classify ads into 14 main policy groups identified by the Comparative Agenda Project (CAP). We compared several state-of-the-art models and chose the one that achieved the best results. During training, we prioritize precision (more than 85% if possible or the highest possible). It is more important to avoid mislabeling ads with the wrong policy category than to miss some advertisements that are related to a policy category.

Analysis of policy-related political ads: We collected and analyzed ads on Meta posted during the 2022 French Presidential Election. We collected political ads from Meta that reached France and were published during the 2022 French presidential election period (Jan 1st, 2022, to June 15, 2022). We use our model to predict policy categories for the collected ads. We observe a significant imbalance in some categories across different demographics groups and regions.

Service to explore French presidential election: We developed a service specifically designed to explore the 2022 French presidential election in detail. The service enables users to access comprehensive statistics about elections, such as how many ads were published, how much money was spent, who are the top advertisers in terms of expenditures, etc. Additionally, the service provides detailed statistics specifically focused on ads mentioning candidates during the first and second rounds and policy-related ads.

Benchmarks for the evaluation of online political ads definition: Numerous definitions of online sponsored political content have been put forward by social media platforms and policymakers, yet there is a lack of metrics available for evaluating them. Therefore, we have formulated four benchmarks to assess the definitions of

online political ads. We evaluated four definitions from different sources (social media platforms and official government documents) based on these benchmarks. The evaluation revealed that all four definitions require further improvement.

Together, these contributions advance the state-of-the-art in analyzing online sponsored political content. We genuinely hope our work will contribute to more effective guidelines for regulations of online political advertising.

7.2 Future work

Furthermore, we have several other ideas for future work, including enhancing our existing work and outlining long-term plans specifically related to online sponsored political content.

7.2.1 Improvement for the detection of policy-related ads

NLP is a rapidly evolving field, constantly progressing. Despite the relatively short time since our article was written, noteworthy advancements have occurred, including the release of models like GPT-3.5 and GPT-4 [67]. These models hold the potential to significantly improve the accuracy of classification of policy-related political ads.

Additionally, Meta has recently made available a dataset [64] containing targeting data for political ads. This dataset will help us understand the reasons behind the unequal distribution of policy-related ads across different demographic groups and different regions.

7.2.2 Analysis of the effect of political ads on users

During my Ph.D. studies, we did not focus on the critical subject of political advertising's effect on users who view it. For future work, we formulated the following questions:

Q_1: *Does the frequency of showing a candidate's ad influence people's voting behavior?*

Q_2: *Does the tone of an ad influence people's voting behavior?*

Q_3: *Do ads that do not mention political candidates and speak only about social issues influence people's voting behavior?*

To answer these questions, our initial task is to develop a tool that tracks whether users receive the ads we send. Another challenge is to create ethical campaigns and ads that will allow us to answer our questions.

Gaining knowledge about the impact of online political ads on users will contribute to a better understanding of which ads require regulation and in what manner.

Acknowledgements

I am deeply grateful to my advisor, Oana Goga, for her exceptional guidance and support during my Ph.D. journey. I am thankful for Oana's significant time and dedication to our projects. She taught me how to conduct research and that you should never trust your data. I sincerely appreciate Oana's patience and support throughout these years.

I would like to thank Kévin Huguenin and Walter Rudametkin for agreeing to review my thesis. I would also like to express my sincere appreciation to Gilles Bastin, Juhi Kulshrestha, and Paolo Frasca for accepting to be members of my thesis committee. It is a pleasure to have you all in my defense committee.

I am grateful to an exceptional researcher and friend, Salim Chouaki. Salim always had answers to my questions, showing me that nothing was impossible. It was a pleasure discussing and working on our research together over these years.

I want to thank Nassim Bouarour, a fantastic colleague and friend, for the great moments we shared during my Ph.D. journey. Our lunch conversations were definitely some of the standout highlights of this experience.

I'm fortunate to have worked with many bright people during this journey. I would like to thank my collaborators, Romaissa Kessi and Maximin Coavoux. I am genuinely proud of the work we accomplished together, and I will always be grateful to them. I would like to express my gratitude to my colleagues from the SLIDE team, Athanasios Andreou, Idir Benouaret, Minh Kha Nguyen, Tinhinane Medjkoune and our great team leader Sihem Amer-Yahia for fostering a friendly and supportive atmosphere within our team.

I want to thank my parents, who encouraged me to do my Ph.D. and pursue my dreams. This Ph.D. journey would not be possible without their support and love. Thank you for always being here for me, even if you are thousands of kilometers away.

Lastly, I want to express my appreciation to the talented scientist and my brother, Dr. Ivan Sosnovik. Ivan has been a source of inspiration for me, and his influence continues to shape my journey. From my earliest days in elementary school to the present,

I am truly grateful for all the assistance and support he has provided throughout the years.

Bibliography

[1] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. Discrimination through optimization: How facebook's ad delivery can lead to biased outcomes. *Proc. ACM Hum.-Comput. Interact.*, 2019.

[2] Jacopo Amidei, Paul Piwek, and Alistair Willis. Rethinking the agreement in human evaluation tasks. In *Proceedings of the 27th International Conference on Computational Linguistics*. Association for Computational Linguistics, 2018.

[3] Athanasios Andreou, Márcio Silva, Fabrício Benevenuto, Oana Goga, Patrick Loiseau, and Alan Mislove. Measuring the facebook advertising ecosystem. In *NDSS*, 2019.

[4] Julia Angwin and Terry Parris Jr. Facebook lets advertisers exclude users by race, 2016. Online available at:https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race.

[5] Ron Artstein and Massimo Poesio. Survey article: Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4), 2008.

[6] Boudhayan Banerjee. Machine learning models for political video advertisement classification. *Creative Components*, 2017.

[7] Bashyakarla V and Hankey S and Macintyre S and Rennó R and Wright G. Personal data: Political persuasion. inside the influence industry. how it works. 2019.

[8] Mohit Baskota. Classification of ad tone in political video advertisements under class imbalance and low data samples. *Creative Components*, 2019.

[9] Tomás Baviera, Javier Sánchez-Junqueras, and Paolo Rosso. Political advertising on social media: Issues sponsored on facebook ads during the 2019 general elections in spain. *Communication & Society.*, 35(3), 2022.

[10] Priyanjana Bengani. As election looms, a network of mysterious "pink slime" local news outlets nearly triples in size. *Columbia Journalism Review*, 2020.

[11] Shaun Bevan. 17Gone Fishing: The Creation of the Comparative Agendas Project Master Codebook. In *Comparative Policy Agendas: Theory, Tools, Data*. Oxford University Press, 2019.

[12] Amber E. Boydstun. Policy agendas topics codebook with media coding addendums, 2014. Online available at: https://comparativeagendas.s3.amazonaws.com/codebookfiles/NYT_Front_Page_Policy_Agendas_Codebook_Updated_with_new_CAP_Codes.pdf.

[13] Leo Breiman. Random forests. *Machine learning*, 45, 2001.

[14] Dafne Calvo, Lorena Cano-Orón, and Tomás Baviera. Global spaces for local politics: An exploratory analysis of facebook ads in spanish election campaigns. *Social Sciences*, 10, 2021.

[15] Arthur Capozzi, Gianmarco De Francisci Morales, Yelena Mejova, Corrado Monti, and André Panisson. The thin ideology of populist advertising on facebook during the 2019 EU elections. In *Proceedings of the ACM Web Conference 2023*, 2023.

[16] Meta Business Help Centre. Core targeting, 2022. Online available at:https://www.facebook.com/business/help/targeting.

[17] Blake Chandlee. Understanding our policies around paid ads, 2019. Online available at: https://newsroom.tiktok.com/en-us/understanding-our-policies-around-paid-ads.

[18] Bruno Coelho, Tobias Lauinger, Laura Edelson, Ian Goldstein, and Damon McCoy. Propaganda política pagada: Exploring u.s. political facebook ads en español. In *Proceedings of the ACM Web Conference 2023*, 2023.

[19] European commission. Public consultation: Commission seeks citizens' views in preparation of new european democracy action plan, 2020. Online available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1352.

[20] European commission. Regulation of the european parliament and of the council on the transparency and targeting of political advertising, 2021. Online

available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1352.

[21] House Intelligence Committee. Exposing russia's effort to sow discord online: The internet research agency and advertisements, 2018. Online available at: https://democrats-intelligence.house.gov/social-media-content/.

[22] Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20, 1995.

[23] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,*. Association for Computational Linguistics, 2019.

[24] Tom Dobber and Claes Vreese. Beyond manifestos: Exploring how political campaigns use misc advertisements to communicate policy information and pledges. *Big Data & Society*, 9, 2022.

[25] Laura Edelson, Tobias Lauinger, and Damon McCoy. A security analysis of the facebook ad library. In *IEEE Symposium on Security and Privacy*, 2020.

[26] US equal employment opportunity commission. Prohibited employment policies/practices, 2023. Online available at: https://www.eeoc.gov/prohibited-employment-policiespractices#:~:text=Terms%20%26%20Conditions%20of%20Employment,)%2C%20disability%20or%20genetic%20information.

[27] European Commission. The digital services act package, 2022. Online available at: https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package.

[28] Facebook. About social issues. Online available at:https://www.facebook.com/business/help/214754279118974.

[29] Facebook. Best practices to choose a category for your facebook page. Online available at:https://www.facebook.com/business/help/376650512904346.

[30] Facebook. Facebook ad library. Online available at: https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=FR&media_type=all.

[31] Facebook. Facebook advertisers' categories. Online available at:https://www.facebook.com/pages/category/.

[32] Facebook. Ads about social issues, elections or politics, 2022. Online available at: https://www.facebook.com/business/help/167836590566506.

[33] Facebook. Become authorised to run ads about social issues, elections or politics, 2022. Online available at: http://bit.ly/379NffN, 2022.

[34] Jennifer Fitzgerald. What does "political" mean to you? *Political Behavior*, 2013.

[35] European Partnership for Democracy. Universal advertising transparency by default, 2020.

[36] Forbes. 2020 political ad spending exploded: Did it work?, 2020. Online available at: https://www.forbes.com/sites/howardhomonoff/2020/12/08/2020-political-ad-spending-exploded-did-it-work/?sh=7e26d2633ce0.

[37] Tommaso Fornaciari, Alexandra Uma, Silviu Paun, Barbara Plank, Dirk Hovy, and Massimo Poesio. Beyond black & white: Leveraging annotator disagreement via soft-label multi-task learning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, 2021.

[38] ERIKA FRANKLIN FOWLER, MICHAEL M. FRANZ, GREGORY J. MARTIN, ZACHARY PESKOWITZ, and TRAVIS N. RIDOUT. Political advertising misc and offline. *American Political Science Review*, 115, 2021.

[39] Abhinav Girdhar. Online vs offline advertising – all you need to know, 2022. Online available at:https://www.appypie.com/online-vs-offline-advertising.

[40] Adina Gitomer, Pavel V Oleinikov, Laura M Baum, Erika Franklin Fowler, and Saray Shai. Geographic impressions in facebook political ads. *Applied Network Science*, 6(1):1–20, 2021.

[41] Google. Political advertising on google, 2018. Online available at: https://adstransparency.google.com/political?political&region=US.

[42] Shloak Gupta, S Bolden, Jay Kachhadia, A Korsunska, and J Stromer-Galley. Polibert: Classifying political social media messages with bert. In *Social, Cultural and Behavioral Modeling (SBP-BRIMS 2020) conference. Washington, DC*, 2020.

[43] Thomas Hansford, Chanita Intawan, and Stephen Nicholson. Snap judgment: Implicit perceptions of a (political) court. *Political Behavior*, 2018.

[44] Libby Hemphill, Annelise Russell, and Angela Schöpke-Gonzalez. What drives u.s. congressional members' policy attention on twitter? *Policy & Internet*, 13, 2020.

[45] Hugging Face. The ai community building the future., 2022. Online available at: https://huggingface.co.

[46] Basileal Imana, Aleksandra Korolova, and John S. Heidemann. Auditing for discrimination in algorithms delivering job ads. *CoRR*, abs/2104.04502, 2021.

[47] Sam Jackson, Feifei Zhang, Olga Boichak, Lauren Bryant, Yingya Li, Jeff Hemsley, Jennifer Stromer-Galley, Bryan Semaan, and Nancy McCracken. Identifying political topics in social media messages: A lexicon-based approach. In *Proceedings of the 8th International Conference on Social Media and Society*. Association for Computing Machinery, 2017.

[48] Noam Scheiber Julia Angwin and Ariana Tobin. Dozens of companies are using facebook to exclude older workers from job ads, 2017. Online available at: https://www.propublica.org/article/facebook-ads-age-discrimination-targeting.

[49] Anja Lambrecht and Catherine Tucker. Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads. *Management Science*, 65(7):2966–2981, 2019.

[50] Hang Le, Loïc Vial, Jibril Frej, Vincent Segonne, Maximin Coavoux, Benjamin Lecouteux, Alexandre Allauzen, Benoit Crabbé, Laurent Besacier, and Didier Schwab. FlauBERT: Unsupervised language model pre-training for French. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. European Language Resources Association, 2020.

[51] legifrance. Loi n° 90-55 du 15 janvier 1990 relative à la limitation des dépenses électorales et à la clarification du financement des activités politiques, 2022. Online available at: https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000000341734.

[52] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *ArXiv*, abs/1907.11692, 2019.

[53] Natasha Lomas. Brexit ad blitz data firm paid by vote leave broke privacy laws, watchdogs find, 2019. https://techcrunch.com/2019/11/27/brexit-ad-blitz-data-firm-paid-by-vote-leave-broke-privacy-laws-watchdogs\-find.

[54] Natasha Lomas. Europe to put forward rules for political ads transparency and beef up its disinformation code next year, 2020. Online available at: http://tcrn.ch/3b4MANV,.

[55] Thomas Louwerse. *Political parties and the democratic mandate. Comparing Collective mandate fulfilment in the United Kingdom and the Netherlands.* PhD thesis, Leiden, the Netherlands, 2011.

[56] Carmen Malena. Working with ngo, 1995. Online available at: https://documents1.worldbank.org/curated/en/814581468739240860/pdf/multi-page.pdf.

[57] Manifesto Project. Manifesto project dataset, 2021. Online available at: https://manifesto-project.wzb.eu/down/data/2021a/codebooks/codebook_MPDataset_MPDS2021a.pdf.

[58] Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamé Seddah, and Benoît Sagot. CamemBERT: a tasty French language model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2020.

[59] Héctor Martínez Alonso, Anders Johannsen, and Barbara Plank. Supersense tagging with inter-annotator disagreement. In *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)*. Association for Computational Linguistics, 2016.

[60] Damon McCoy. Online political ads transparency project, 2021. Online available at: https://engineering.nyu.edu/research/online-political-ads-transparency.

[61] Sam Meredith. Facebook-cambridge analytica: A timeline of the data hijacking scandal, 2018. Online available at: https://www.cnbc.com/2018/04/10/facebook-cambridge-analytica-a-timeline-of-the-data-hijacking-scandal.html.

[62] Meta. Ad library report, 2019. Online available at: https://www.facebook.com/ads/library/report/?source=archive-landing-page&country=FR.

[63] Meta. Doing more to protect against discrimination in housing, employment and credit advertising, 2019. Online available at: https://about.fb.com/news/2019/03/protecting-against-discrimination-in-ads/.

[64] Meta. Ad targeting dataset, 2023. Online available at: https://developers.facebook.com/docs/fort-ads-targeting-dataset/overview.

[65] Giovanni Moretti, Rachele Sprugnoli, and Sara Tonelli. Digging in the dirt: Extracting keyphrases from texts with kd. 2015.

[66] NYU Cybersecurity for Democracy. Explore facebook and instagram political ads, 2022. Online available at: https://adobservatory.org/?search_by=topic.

[67] OpenAI. Models, 2023. Online available at: https://platform.openai.com/docs/models.

[68] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010.

[69] European Parliament. Committee on culture and education, 2022. Online available at: https://www.europarl.europa.eu/doceo/document/CULT-AD-735573_EN.pdf.

[70] Barbara Plank, Dirk Hovy, and Anders Søgaard. Learning part-of-speech taggers with inter-annotator agreement loss. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 2014.

[71] Plotly. Dash python user guide, 2023. Online available at: https://dash.plotly.com.

[72] Victor Le Pochat, Laura Edelson, Tom Van Goethem, Wouter Joosen, Damon McCoy, and Tobias Lauinger. An audit of facebook's political ad policy enforcement. In *31st USENIX Security Symposium*. USENIX Association, 2022.

[73] Prolific. Prolific, 2022. Online available at: https://www.prolific.co.

[74] ProPublica. Political advertisements from facebook. Online available at: https://www.propublica.org/datastore/dataset/political-advertisements-from-facebook.

[75] ProPublica. Facebook political ad collector, 2020. Online available at: https://projects.propublica.org/facebook-ads/?lang=en-US.

[76] Qualitrix. Qualtricsxm, 2022. Online available at: https://www.qualtrics.com.

[77] Filipe N. Ribeiro, Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnnatan Messias, Fabricio Benevenuto, Oana Goga, Krishna P. Gummadi, and Elissa M. Redmiles. On microtargeting socially divisive ads: A case study of russia-linked ad campaigns on facebook. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 2019.

[78] Annelise Russell. Gendered priorities? policy communication in the u.s. senate. *Congress & the Presidency*, 48, 2021.

[79] Giovanni Sartori. What is "politics". *Political Theory*, 1973.

[80] Scikit-learn. Scikit-learn: Machine learning in python, 2007. Online available at: https://scikit-learn.org/stable/,.

[81] Congressional Research Service. Online political advertising: Disclaimers and policy issues, 2019. Online available at: https://crsreports.congress.gov/product/pdf/IF/IF10758.

[82] Márcio Silva, Lucas Santos de Oliveira, Athanasios Andreou, Pedro Olmo Vaz de Melo, Oana Goga, and Fabrício Benevenuto. Facebook ads monitor: An independent auditing system for political ads on facebook. In *TheWebConf*, 2020.

[83] Vera Sosnovik and Oana Goga. Understanding the complexity of detecting political ads. In *Proceedings of the Web Conference 2021*, page 2002–2013, 2021.

[84] Vera Sosnovik and Oana Goga. How to define political ads? In *Proceedings of the 3rd International CoNEXT Student Workshop*, 2022.

[85] Vera Sosnovik, Romaissa Kessi, Maximin Coavoux, and Oana Goga. On detecting policy-related political ads: An exploratory analysis of meta ads in 2022 french election. In *Proceedings of the ACM Web Conference 2023*, 2023.

[86] Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe Nunes Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. Potential for discrimination in online targeted advertising. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 2018.

[87] Ed Summers. irads, 2018. Online available at: https://github.com/umd-mith/irads/blob/master/README.md.

[88] Daniel Susser, Beate Roessler, and Helen Nissenbaum. Online manipulation: Hidden influences in a digital world. *SSRN Electronic Journal*, 2018.

[89] Zhanna Terechshenko, Fridolin Linder, Vishakh Padmakumar, Fengyuan Liu, Jonathan Nagler, Joshua Tucker, and Richard Bonneau. A comparison of methods in political science text classification: Transfer learning language models for politics. *SSRN Electronic Journal*, 2020.

[90] TextBlob. Textblob: Simplified text processing. Online available at: https://textblob.readthedocs.io/en/dev/.

[91] the Electoral Commission. Know who is paying for misc political ads, 2022. Online available at: https://www.electoralcommission.org.uk/i-am-a/voter/misc-campaigning/know-who-paying-misc-political-ads.

[92] Robert Thomson. *Resolving controversy in the European Union: Legislative decision-making before and after enlargement*. Cambridge University Press, 2011.

[93] Robert Thomson. 340Parties' Election Manifestos and Public Policies. In *The Oxford Handbook of Political Representation in Liberal Democracies*. Oxford University Press, 2020.

[94] Ariana Tobin and Jeremy B. Merrill. Facebook is letting job advertisers target only men, 2018. Online available at: https://www.propublica.org/article/facebook-is-letting-job-advertisers-target-only-men.

[95] Twitter. Political content, 2023. Online available at: https://business.twitter.com/en/help/ads-policies/ads-content-policies/political-content.html.

[96] Mark Warren. What is political? *Journal of Theoretical Politics*, 1999.