On Meta's Political Ad Policy Enforcement: An analysis of Coordinated Campaigns & Pro-Russian Propaganda

Paul Bouchaud

On Meta's Political Ad Policy Enforcement: An analysis of Coordinated Campaigns & Pro-Russian Propaganda

Paul Bouchaud
Center for Social Analysis and Mathematics (EHESS)
Complex Systems Institute of Paris Île-de-France (CNRS)
AI Forensics
Paris, France
paul.bouchaud@ehess.fr

ABSTRACT

This study assesses Meta's enforcement of its political advertising policy across 16 EU countries using the Meta Ad Library. Our analysis finds Meta's moderation imprecise, characterized by a 60% false-positive rate, a false-negative rate of 95.2%, with notable disparities across countries. Within undeclared political ads, we identify coordinated advertising campaigns, including cross-country investment scams reaching a total of 128 million accounts in January-February 2024. Alongside, a network disseminating pro-Russian propaganda, part of the so-called Doppelganger operation, accumulated a total reach of 38 million accounts in France and Germany between August 2023 and March 2024. Despite documented activities, among the 3,826 Facebook pages involved in disseminating propaganda ads undermining support for Ukraine and institutional support in the EU, less than 20% were moderated by Meta, after they had already been shown by Meta to users at least 2.6 million times. Our findings underscore significant shortcomings in Meta's moderation of political advertisements, emphasizing the need for enhanced enforcement measures, particularly in light of the upcoming elections.

1 INTRODUCTION

Advertising fuels a sizable part of the web's economy and plays an increasingly important role in political elections. Digital platforms host marketplaces where advertisers vie for user attention through ads, with intermediaries like Google, Facebook[1], and Amazon leveraging user data for targeted advertising [22, 42]. In the aftermath of the 2016 United States presidential election, marked by the abuse of targeted advertising on Facebook, social media platforms emerged as battlegrounds for information and attention wars. The Russian Internet Research Agency run prior to 2016 U.S. elections, advertisement, through Facebook's platforms, on polarizing topics targeting vulnerable sub-populations [51]

In response to these revelations and heightened scrutiny, Facebook revised, in 2018, its Terms of Service policy to restrict the dissemination of political ads solely to advertisers residing in the same country as their targeted audience [33]. Additionally, all political advertisements are mandated to carry conspicuous labels, prominently featuring a "Paid for by" disclosure from the advertiser at the forefront of the ad. Furthermore, Facebook introduced an Ad Library [32, 36], covering European Union members by May 2019 [34], enabling users to access a repository of ads categorized as political, along with details such as the advertiser's identity, the number of impressions and targeted demographics. In August 2023, in compliance with the European Union's Digital Services Act, Meta expanded its Ad Library to archive *all* advertisements targeting individuals within the EU [14]. Nevertheless, the amount spend by advertisers, the number of impressions and other metadata are only made public for political ads [2].

Despite enhanced transparency, criticism from civil society organizations and academics points to significant shortcomings in Meta's advertising ecosystem [56] in terms of transparency and accountability. [1, 21] reveal notable lapses in data integrity, with ads being removed from the library, despite Meta's commitment to archive political advertisements for seven years. Furthermore, Meta's disclosure of targeting information remains limited to basic demographic details, which, considering the wide range of targeting criterion offered by Meta to advertisers, offering little insight into the ad delivery systems [25, 51, 55]. Additionally, Meta's reliance on voluntary self-declaration for political ads presents a critical loophole, allowing dishonest actors to evade scrutiny by omitting political labels from their advertisements. [53] identified undeclared political ads during Brazil's 2018 presidential election were absent from the Ad Library due to advertisers' failure to categorize them as such. Similarly, advertisements associated with the Doppelganger pro-Russian disinformation campaign, identified by civil society organizations [12, 23, 26, 45, 48] and the French service combating foreign digital interference [58], were not self-declared as political.

In addition, to self-declaration, Meta claims that it rigorously review each advertisement against its Advertising Standards [5], in particular its policy on ads about *social issues, elections or politics*. Ads lacking appropriate disclaimers but flagged by Meta as containing such political content are rejected during initial review. Active ads may be flagged by automated systems or reported by users, if found to violate Meta's policy due to the absence of a disclaimer, they are disapproved and included in the Ad Library. Importantly, it should be noted that the Meta Ad Library does not encompass ads undergoing pre-launch moderation. In 2022, Le Pochat et al. performed an Audit of Meta's Political Ad Policy Enforcement, highlighting an imprecise moderation where 61% more ads are missed than are moderated worldwide, and 55% of ads moderated by in the U.S. Meta as political are in fact non-political [27]. Previously, in

[1] In this paper, "Meta" is used to refer to the company after 2021, while "Facebook" is used to denote the social media platform.

[2] In this paper, we use "political ads" to denote advertisements falling within Meta's policy, encompassing "ads about social issues, elections, or politics."

2019, Edelson et al. detected advertisers engaged in undeclared coordinated activity spending millions of dollar on political advertising [18].

Leveraging comprehensive access to the Meta Ad library granted by the Digital Services Acts, we conducted an assessment, across 16 EU countries, of Meta's Political Ad Policy Enforcement. We estimate the prevalence of undeclared political advertising on Meta and identify coordinated campaigns operating without proper declaration. Consistent with previous research, our results indicate a significant level of false positives in Meta's moderation efforts, with 60% of moderated ads being inaccurately flagged. Moreover, our analysis highlights a notable deficiency in Meta's ability to identify undeclared political ads, with an estimate 95.8% remaining undetected. Within this subset of undeclared political ads, we uncovered cross-national investment scams campaigns, reaching over a hundred millions account in January-February 2024. Additionally, we detect, between August 2023 and March 2024, a network large of 3 826 pages disseminating pro-Russian propaganda, having reached 38.0 million accounts in France and Germany, with less than 20% of the ads being moderated by Meta. Our findings emphasize the critical need for proactive measures to uphold transparency, accountability, and democratic integrity, in the context of forthcoming elections and ongoing geopolitical conflicts.

2 POLITICAL ADVERTISING

In this study, we aim to evaluate Meta's enforcement of its political ads policy, without delving into a broader discussion on what constitutes political ads per se. As investigated by Sosnovik and Goga, the definition of political ads is open to interpretation, leading to significant discrepancies among ad platforms, the general public, and advertisers regarding what qualifies as political advertising, particularly concerning advertisements addressing social issues. [54] indicates that volunteers tend to not classify ads from NGOs and charities as political, while advertisers do categorize them as such. Conversely, advertisers tend to not classify social issue ads as political, contrasting with volunteers. While enhanced guidelines may help mitigate some of this discrepancy, many advertisements addressing societal and humanitarian issues inherently pose challenges for precise labeling.

The new European Union regulation on transparency and targeting of political advertising [38], considered ads as political if they either i) are created by, for, or on behalf of a political actor, unless solely of a private or commercial nature; or ii) are intended and designed to influence the outcome of an election or referendum, voter behavior, or a legislative or regulatory process. Exceptions include communications promoting or discussing methods for participating in elections or referendums, including the announcement of candidacies or the referendum question.

Adopting a broader definition, Meta considered ads as related to social issues, elections, or politics —hereinafter referred to as *political ads* for brevity- if i) "*made by, on behalf of or about a candidate for public office, a political figure, a political party, a political action committee or advocates for the outcome of an election to public office; ii) About any election, referendum or ballot initiative, including "get out the vote" or election information campaigns; iii) About any social issue in any place where the ad is being run is regulated as political advertising*" [4]. In the European Union, Meta considers the following top-level social issues when reviewing ads: *Civil and Social Rights, Crime, Economy, Environmental politics, Health, Immigration, Political values and governance, Security and Foreign Policy* [6].

To help prospective advertisers, Meta provides examples of ads, related to social issues, requiring or not, to be declared as political [11]. For instance, "It's critical for trans people to be given their own voice" is deemed political while "Civil rights exhibition opens on Monday" is not. Similarly, Meta provide examples of ads selling products or services considered or not as political, for instance "Save on the fridge that shrinks your carbon footprint – now until December." is not required to be labeled as political while "Our company now operates on 100% renewable energy. We must protect the environment." does [8].

In order to run such political advertisements on Meta, advertisers must authenticate their accounts by presenting proof of identity [10], and are limited to sharing political ads exclusively within their country of residence [9]. Moreover, advertisers are also mandated to disclose the funding source, the disclaimer "Paid for by" is then prominently displayed atop the ad frame.

3 METHODS

3.1 Data Collection

3.1.1 Meta Ad Library. Since August, 17th 2023, Meta made public an Ad Library, accessible through a web portal and an API [36], archiving "all ads that target people in the EU". In this work, we leverage a comprehensive data collection [29], performed through the Meta Ad Library API, of all ads ran in 16 European countries between August, 17th 2023 and February, 29th 2024. Every hour, the 200k newest ads were crawled [29]. Every week the ads metadata, such as their reach or political status, were updated. For each advertisement, we collected the following metadata: *the displayed text, link's captions, delivery start and end dates, demographic information concerning the accounts reached within the EU by the advertisement, the estimated overall reach, and the age, gender, locations targeted by advertisers*. In cases where advertisements are identified as political, whether by self-declaration or Meta's moderation, Meta discloses various metrics, including *the number of impressions, and the budget expenditure*, within a defined range.

Meta defines "reach" as the number of accounts that have viewed the ads at least once, while "impressions" refer to the number of times the ads have been displayed on users' screens [3]. Reach differs from impressions in that it may include multiple views of the ads by the same accounts. These metrics are directly impacted by advertisers' bidding, budget allocations, and audience targeting criteria.

Meta claims to consider in its ads moderation system, the text, images, videos and landing pages associated to the ads [6]. Nevertheless, one does not have a straightforward access, through the Meta Ad Library API, to potential media associated to ads. Hence, we solely rely in this work on text, filtering out ads with less than 10 characters. Future work may consider media, as performed in [53], contingent upon enhancements to the Meta Ad Library API, or through extensive scraping of the Meta Ad library web portal and at important computational cost.

We consider in this study the 10 most used languages for advertisement in the European Union between August and December 2023 and the 10 most used languages for ads declared as political in the EU. When the language of an ad was missing from the metadata disclosed by Meta, we detected it using FastText [24]. The set of 13 languages selected for this study is then: Bulgarian, Dutch, English, French, German, Greek, Hungarian, Italian, Polish, Portuguese, Romanian, Slovak, Spanish, Swedish; accounting for over 90% of all advertisement running in the EU during the considered timeframe. Furthermore, we will restrict our analysis to ads targeting a set of 16 countries, having for official languages one of the selected languages, and considered only ads written in the language of the country. Specifically, we considered the following EU countries: Austria (German), Belgium (Dutch, French & German), Bulgaria (Bulgarian), France (French), Germany (German), Greece (Greek), Hungary (Hungarian), Ireland (English), Italy (Italian), Netherlands (Dutch), Poland (Polish), Portugal (Portuguese), Romania (Romanian), Slovakia (Slovak), Spain (Spanish) and Sweden (Swedish).

3.1.2 Training Dataset. We seek to train a machine learning model capable of classifying advertisements, based on their textual content, as either requiring or not requiring to be declared as political under Meta's guidelines. The training process will be performed on ads ran in the languages and countries listed above, spanning from August 17th, 2023, to December 31st, 2023. Ads ran in January and February 2024 will be considered for inference purposes.

The training dataset was curated following Meta's guidelines. Initially, we included 24 political ads and 20 non-political examples provided by Meta to advertisers, which were translated into 14 languages. While these examples offer some guidance, they tend to be generic and may not fully represent the typical format of most ads. Subsequently, we conducted a semantic search within each language's pool of collected ads to identify those that closely resembled Meta's examples. This search was performed via the "paraphrase-multilingual-MiniLM-L12-v2" multilingual model [46]. As a result, we obtained a labeled and balanced dataset consisting of 560 ads.

We augmented the training dataset by including ads that, for each country, mentioned the name of the head of state, the main political parties within the country's respective legislative body, and their leaders. This addition was prompted by Meta's specification in its Business Help center, which stipulates that ads will be considered political if they feature the "name of a political figure, politician, or candidate for public office" [6]. Furthermore, we conducted searches for ads containing specific political keywords: "elections," "minister," "president," "mayor," "law," and "regulation." This approach aligns with Meta's policy, which defines political ads as those advocating for "a change to law or policy, for or against legislation such as regulations, decrees, executive orders," as well as encompassing "grassroots advocacy where there is a call to action to a person to contact an elected official or governing body to take a specific step or to sign a petition aimed at an elected official or governing body" [6]. Overall, we expanded our training dataset by an additional 300 annotated ads through this process.

To train our model to adhere to Meta's moderation guidelines and identify political advertisements undeclared by advertisers, we relied on ads moderated by Meta. However, as emphasized in previous studies, caution is warranted in considering Meta's moderation as the gold standard due to a high false-positive rate [18, 27]. Notably, we observed instances where ads moderated by Meta were largely unrelated to social issues. Consequently, we manually labeled an additional 800 ads, 50 randomly selected ads in each country, from those moderated by Meta. Our annotation process systematically referred to Meta's official moderation guidelines and provided examples to ensure accuracy. Furthermore, we curated a set of 1 120 non-political ads, with 70 selected from each country, from the pool of ads not identified as political by advertisers and not moderated by Meta, primarily featuring commercial products. In total, we annotated over 3.3k ads, balanced across 14 languages, to train our model.

3.1.3 Evaluation Dataset. To assess Meta's moderation effectiveness and calibrate our model scores, we conducted additional manual annotations. In particular, for each country, we annotated i) a sample of 100 advertisements, 10 from each model score decile, ii) 100 advertisements categorized as non-political by their advertisers, iii) 100 advertisements moderated by Meta, and iv) 100 advertisements not declared as political and not moderated by Meta. Each advertisement underwent triple coding, once by a human reviewer, once through the GPT-3.5 Instruct model [39, 40], finally discrepancies were settled by independent human reviewers. Annotators aimed to adhere closely to Meta's guidelines. When uncertain, we cross-referenced the Meta Ad Library to determine if, multiple, similar advertisements had been moderated by Meta. When doubt persisted, edge cases were settled for Meta's assessment i.e. declared or moderated ads were labeled as political and non-declared or non-moderated ads as non-political. Our subsequent results are then conservative estimates in favor of Meta.

3.2 Model Training

3.2.1 Pre-processing. The dataset was pre-processed by removing URLs, special characters, and converting emojis to their CLDR names [13]. Furthermore, the ads text was truncated to the first 500 characters to standardize text length.

3.2.2 Training. Following the curation of the training dataset, we employed a two-step training architecture as described in [57], particularly efficient with small labeled datasets. Initially, we fine-tuned a multilingual sentence-transformer, "paraphrase-multilingual-MiniLM-L12-v2" [46], utilizing contrastive learning. The model was selected due to its pre-training across all 14 languages of interest and its balanced performance in semantic similarity tasks relative to computational resources. Subsequently, a logistic regression model was trained as a classification head over the fine-tuned embeddings. Evaluation on a 10% holdout dataset resulted in an F1-score of 0.88 and an AUC of 0.95 for the trained model.

For comparison, we implemented the supervised learning algorithms emloyed in [53] for detecting political ads within a monolingual context. Specifically, a Naive Bayes classifier utilizing Hashing Vectorizer for feature extraction achieved an F1-score of 0.64 and an AUC of 0.73.

3.3 Evaluation of Meta's Enforcement of Political Ads Policy

Our objective is to assess the effectiveness of Meta's enforcement of its political ads policy. We will provide an overview of the overall statistics regarding the number of ads published, self-declared and moderated by Meta. Subsequently, we will evaluate Meta's moderation false-positive rate through manual annotation, i.e. the fraction of ads classified by Meta as political despite not falling under its guidelines. Leveraging the ads flagged as political by our model, we will estimate the proportion of undeclared political ads and Meta's moderation recall.

3.3.1 Evaluation of Ads Featuring Heads of State. As discussed in by Sosnovik and Goga, the definition of ads *about social issues, elections, or politics* is broad and subject to interpretation [54]. Our initial examination focuses on a subset of ads that are unequivocally political, namely ads, ran in Germany, France, and Italy from August 2023 to February 2024, which feature the respective heads of state: Olaf Scholz, Emmanuel Macron, and Giorgia Meloni. This dataset is curated to solely encompass ads directly associated with political matters, all of which fall within the scope of Meta's policy mandating declaration of ads containing "the name of a political figure" [6]. Subsequently, we analyze the proportion of ads declared by their advertisers as political, alongside the fraction of undeclared ads having been moderated by Meta. Meta may exempt specific news providers from declaring their political ads [7]. However, as the list of such exempted providers is not, to our knowledge, publicly available, we report the recall rates for moderation and declaration of ads featuring heads of state, with and without excluding ads originating from Facebook pages associated with news providers, manually curated under a broad definition encompassing both national and local press.

3.3.2 Political and Social Issue Ads. Expanding our analysis to the full scope of ads classified by Meta as relating to *social issues, elections, or politics*, we leveraged our trained model to score the 29.5 million ads published in January and February 2024 across the specified countries and languages. Subsequently, we computed the fraction of undeclared political ads moderated by Meta, i.e. the ratio of correctly moderated ads to the total number of undeclared political ads estimated by our model. Furthermore, among the ads flagged as political by our model, we calculated the proportion that were declared as political by their advertisers. As emphasized above, the manual annotation is such that our results are conservative estimates in favor of Meta.

3.3.3 Repeated Violations. Finally, we investigate Meta's claim that repeated failure to disclose ads as political may result in penalties against the advertiser [2]. Specifically, we identify pages having been repeatedly moderated, at least one ads getting moderated per week on average between August and February 2024. We then assess the proportion of these pages that remain active and continue to advertise on Meta.

3.4 Coordination

3.4.1 Detection. Having scored the entire set of ads instead of a limited sample, we can explore the set of undeclared political ads. In particular, we focus in this work on the detection of coordinated behaviors, in the form of apparently independent pages publishing highly similar ads. To tackle this task, [18] successfully used a simhash algorithm [31], equivalent to measuring the Hamming distance between two ads texts. However, such method will fail to detect text having been translated or reworded. With the ever-decreasing cost of translation tools and advent of generative AI, such content duplication method has made it prevalent in social networks [19]. We then adapt the 3Δ detection method of *copy-pasta*, *rewording*, and *translation* introduced by Richard et al. [52] for social media content. *Copy-pasta* refers to duplicating content with minor modifications. The 3Δ method compares pairs of text message along three specific dimensions, each seeking to detect an artificial content amplification technique, namely: semantic meaning, wording (characterized by graphemes), and the language employed in the message [52]. Pairwise distances within each dimension are computed to identify messages that exhibit close proximity.

In order to alleviate the computational burden associated to pairwise comparison over a large corpus, we adapted the algorithm introduced in [52], leveraging Faiss [17] for efficient similarity search. The detection of coordinated behaviors is then performed following the algorithm prescribed in Alg.1. In particular, ads text embedding vectors are computed with the "paraphrase-multilingual-MiniLM-L12-v2" [46] model. It produces aligned vectors for similar inputs in different languages [47] and has been trained for semantic similarity search, making it particularly relevant for the present use. Then, we search for ads having a cosine similarity distance between their embedding vectors smaller than τ_s, empowered by Faiss efficient indexing. For each pair of semantically close ads, we compare their languages. If the two ads are in different languages, then we consider them as *translations*. If not, we compute the Levenshtein distance between the two ads' text, normalized by texts length. Below a threshold τ_g, we consider ads as *copy-pasta*, above as *rewording*. The threshold τ_s was determined, on a corpus of translated pairs of text, and τ_g was fixed by expert knowledge on pairs of text, differentiating *copy-pasta* from *rewording*.

We can analyze the pairs of ads identified by the 3Δ method as a network, where nodes are ads and edges are duplication strategies. By identifying connected components in the network, we detect set of ads linked to each other by being a *rewording*, a *translation* or a *copy-pasta* from one another.

Firstly, in an agnostic approach we seek to detect coordination among the set of undeclared political ads, published in January and February 2024. Specifically, we consider the set of 320k ads, not declared as political, over the 16 countries of interest, that our model score above a threshold associated to a recall of 90%. Following this initial exploration, and considering previous works [12, 48, 58], we collected, and performed the coordination detection analysis on, the set of 5 288 and 6 949 ads having targeted France and Germany between August 17th, 2023 and March 31th, 2024, mentioning Ukraine; matching the case insensitive regular expression `\bukrain\w*`.

3.4.2 Topic Modelling. To delve into the content of ads published within coordinated campaigns, we conducted topic modeling. Despite recent progress in neural natural language processing, we

Algorithm 1 Coordination Detection, adapted from [52]

Require: τ_s: Semantic distance threshold
Require: τ_g: Grapheme distance threshold
Index ads text embedding with Faiss
for each *Ad* in *Corpus* **do**
  Let *Neighbors* be an empty list
  Search for neighbors of *Ad* within τ_s using the Faiss index
  Add found neighbors to *Neighbors*
  for each *Neighbor* in *Neighbors* **do**
    if lang(*Ad*) ≠ lang(*Neighbor*) **then**
      return "Translation"
    else
      if $\Delta_{\text{grapheme}}(Ad, Neighbor) < \tau_g$ **then**
        return "Copy-Pasta"
      else
        return "Rewording"
      end if
    end if
  end for
end for

opted for a classical approach that combines TF-IDF (Term Frequency-Inverse Document Frequency) with Non-Negative Matrix Factorization (NMF). This decision was motivated by the simplicity, efficiency, and robustness of TF-IDF/NMF, as emphasized in [59].

NMF decomposes the term-document matrix generated through TF-IDF into two non-negative matrices: one representing terms and topics, and the other representing topics and documents. Such decomposition facilitates straightforward interpretation [28]. The number of topics is chosen to maximize topic coherence, assessed through the *word2vec*-based metric introduced in [41].

4 RESULTS

4.1 Meta's Political Ads Policy Enforcement

4.1.1 Ads Featuring Heads of State. Table 1 displays the proportions of ads containing the names of the German, French, and Italian heads of state that were declared as political by advertisers, alongside the fractions of undeclared ads moderated by Meta under its political ads policy. Our analysis indicates that Meta's moderation of ads falling within its political ads policy remains limited across all three countries, regardless of the potential exemptions for ads from news providers. Notably, there are variations in Meta's moderation recall among the countries examined. For instance, in France, only 16.3% of undeclared ads featuring the name of President Emmanuel Macron were moderated, compared to 42.3% of undeclared ads containing the name of Chancellor Olaf Scholz in Germany. Additionally, we observe differences in advertiser compliance with self-declaration requirements for political ads. Specifically, 87.9% of ads including the name of Prime Minister Giorgia Meloni in Italy were declared as political, while this percentage decreases to 81.6% for ads featuring President Macron in France.

4.1.2 Political and Social Issue Ads. In Table 2, alongside the general statistics regarding the daily volume of published ads, we present the precision and recall associated with advertisers' self-declaration of political ads and Meta's moderation. Among the 62k, 50k, and 46k ads launched daily in Germany, Italy, and France, Meta moderates an average of 34.0, 34.4, and 30.1 ads. Political ads, self-declared by advertisers, constitute 0.53% of all advertisements published across the 16 countries, with notable variations between countries, from 0.18% in Portugal to 1.45% in Hungary. On average, over the 16 countries of interest and weighted by the daily number of ads declared as political, only 57.4% of ads declared as political required such disclosure under Meta's guidelines. Similarly, 60.4% of ads moderated by Meta did not align with Meta's criteria for political advertising (country-wise results averaged, weighted by the number of moderated ads). Through our model, we establish a conservative lower bound for the number of undeclared political ads launched daily in each of the 16 countries. Among ads identified as political by our model, only 33.9% were declared by their advertisers, exhibiting significant variations across countries, ranging from 49.6% in the Netherlands to 10% in Portugal.

Additionally, Table 2 presents the proportion of undeclared political ads moderated by Meta. Ireland, the sole English-speaking country in our study, exhibits the highest moderation recall, with 22.8% of undeclared ads being moderated by Meta. Conversely, in Austria, only 3.4% of the undeclared ads identified as political by our model were moderated. Across the 16 countries, weighting Meta's recall country-wise by the number of undeclared political ads, Meta's moderation recall stands at 4.8%.

4.1.3 Repeated Violations. Over the period August 17th, 2023 to February 29th, 2024, we identified 92 advertising pages that had been moderated at least 28 times, corresponding to an average of one moderated ad per week. Within this dataset, the most moderated pages accumulated 315 moderated ads. Of the 92 pages that repeatedly failed to disclose the political nature of their ads, as assessed by Meta, 81 were still active as of March 29th, 2024, with 73 of them running ads in March 2024; only 10 pages were deleted. Due to lack of information, it is unclear whether these pages were taken down by Meta or simply deleted by their owners.

Table 1: Number of ads, Moderation and Declaration Recalls by Country, evaluated on ads containing the name of the head of state, with and without filtering ads published by news providers

Country	Metric	With Media	Without Media
Germany	Number of Ads	306	219
	Moderation Recall	53.5%	42.3%
	Declaration Recall	85.9%	88.1%
France	Number of Ads	465	234
	Moderation Recall	8.7%	16.3%
	Declaration Recall	63.0%	81.6%
Italy	Number of Ads	1460	1409
	Moderation Recall	38.1%	37.4%
	Declaration Recall	87.9%	87.9%

Metric	Overall	Austria	Belgium	Bulgaria	France	Germany	Greece	Hungary	Ireland	Italy	Netherlands	Poland	Portugal	Romania	Slovakia	Spain	Sweden
Ads Launch Daily	**446 548**	25 938	33 565	7 408	46 561	62 742	9 489	15 440	76 221	50 451	24 896	34 871	18 220	17 685	7 742	38 287	13 736
Ads Declared as Political Daily	**2 372**	141.0	148.3	18.2	110.0	245.9	104.1	228.2	35.2	438.8	93.0	232.4	33.4	210.2	112.2	105.1	123.7
Ads Moderated Daily	**228**	8.8	9.6	4.6	30.1	34.0	4.3	7.9	20.5	34.2	7.9	20.9	7.8	8.6	4.7	24.2	9.2
Declaration Precision (%)	**57.4***	70.6	71.6	55.9	41.9	66.7	45.4	52.6	47.6	51.1	57.3	52.0	56.5	73.5	61.0	51.7	50.6
Moderation Precision (%)	**39.6***	33.3	34.3	38.5	28.6	49.4	42.7	48.0	41.2	40.3	32.9	45.8	41.2	48.4	48.9	29.3	42.2
Undeclared Political Ads Daily	**1 356.9**	83.8	73.0	21.8	162.4	135.9	40.1	93.9	37.1	236.7	24.3	90.6	51.7	98.4	37.6	126.7	42.9
Declaration Recall (%)	**33.9**‡	37.1	42.8	20.4	13.7	36.5	39.5	36.5	12.4	37.4	49.6	30.7	10.0	48.3	50.2	14.5	44.0
Moderation Recall (%)	**4.8**†	3.4	4.4	7.5	5.3	12.4	4.4	3.9	22.8	5.8	10.0	10.6	6.1	4.2	5.6	5.6	8.9

Table 2: Daily advertisement statistics across countries. Moderation and Declaration precision were evaluated through manual annotation, and Moderation and Declaration recall using our model. The "Overall" columns present aggregated metrics across 16 countries, accounting for cross-country advertisements. Precisions and recalls are weighted averages per country, with weights corresponding to the number of declared ads (⋆), moderated ads (∗), political ads (†), and undeclared political ads (‡). Due to the conservative nature of our model, the count of undeclared political ads serves as a lower bound, resulting in inconsistencies when aggregating averages. Analysis was conducted on advertisements from January and February 2024, in the respective official languages of the 16 countries, meeting a 10-character length requirement. Please refer to the data collection section for more information.

4.2 Ads Duplication & Coordinated Campaign

The network defined by ads linked to each other by duplication strategies, identified via the 3Δ algorithm, is large of 248k nodes and 24.7 millions edges. In particular, we identify, 443k pairs of translated ads, 1.91 millions pairs of reworded ads and 22.3 millions pairs of copy pasta, including 20.5 millions exact duplication. The network contains 36 276 connected components, the median number of nodes/ads by connected components equals 2, with only 3 292 connected components larger than 10 ads. 90.3% of the connected components are made of ads published by the same page. Indeed, an advertiser may publish multiples variation of the same ads, or published the exact same ads multiples times when one expire or changing the targeting options. We observe that each unique advertisement text is, on average, published 2.4 times by the same page, with 5% of them being published more than 6.1 times by the same page.

4.2.1 Cross-country campaigns. Over the 4 721 connected components made of ads targeting more than one country, 91.0% of them targeted only two. Such two countries campaign, tend to target countries with the same language, 92.4% of such campaigns are targeting either: Austria & Germany, Belgium & France, Belgium & Netherlands, or Belgium & Germany. 90.9 % of two countries components are made of ads run by a single page.

When sorting the connected components by number of country targeted by the ads, we identify multi-country campaigns. For instance, we uncover a 17 ads campaign, seeking testimony of survivors of sexual violence in childhood in order to strengthen child protection measures and rights of victims. The campaign is translated and targets 11 of the 16 countries considered in this paper, have reached 749k accounts, without being declared as political nor being moderated by Meta as requiring to be declared as being about social issues and politics. Similarly we identify three campaigns launch by Madeira tourism office praising environmental, economic and cultural sustainability, the 30 ads total a reach of 1.6 millions accounts in 8 out of 16 countries. Falling under Meta political ads policy, 11 ads were moderated, after having reached a total of 11 844 accounts.

4.2.2 Scam campaigns. Interested in detecting coordination between multiple pages, we sort the connected components by number of unique pages involved and filtered out one-page-component, resulting in 3 498 connected components. The largest component consists of 1,219 ads targeting France, comprising three unique text variations, launched continuously over the observation window. These ads were published by 266 unique pages and have collectively reached a total of 36 million accounts. The ads follow the pattern "The host of [TV Show] [Name of the host] described [Name of a journalist, celebrity or political figure] as irresponsible and declared live that financial information of such magnitude could shake the foundations of [country] society" and redirect users toward 103 different landing pages, we display such an adi n Figure 6 in Appendix. The landing pages of, now inactive, ads project the facade of legitimate e-commerce websites. However, scrutinizing the web page code reveal their counterfeit nature, being front-end replicas of legitimate e-commerce platforms. Searching for active ads with the same textual content, reveals that the landing pages of active ads redirect users to forged media websites posing as reputable newspapers and media organizations. Within these counterfeit news platforms, users are being displayed interviews featuring celebrities endorsing schemes promising wealth through the promotion of cryptocurrency programs, accompanied by redirection links to said investment platforms. The fraudulent nature of such platform as been extensively investigated in [12, 20, 45, 48].

The subsequent largest components follow the same pattern, considering the additional textual pattern "Financial information of this magnitude can shake the foundations of [country] society" differing only by the language and targeted country. Overall, we identify 82 connected components matching these patterns, for a total of 8 232 ads, published by 971 unique pages and targeting 10 out of 16 countries of interest: Austria, Belgium, France, Germany, Greece, Hungary, Italy, Netherlands, Portugal and Spain. Between January and February 2024, this campaign of scam-ads accumulated a total reach of 128.2 millions accounts, mainly in France, Italy and Spain; we report statistics of this network, broken down by country, in Appendix. Among those ads, none were declared as related to social issues by their author, despite referring to financial institutions, and less than 0.1% of them were moderated by Meta. We inspected the Meta Ad Library web portal and find moderation decisions in

2023 of ads following the above identified pattern, confirming the alignment between Meta guidelines and our detection.

4.2.3 Political Campaigns. After investments scam-ads, the subsequent largest connected components, in terms of the number of unique pages, relay political messages. Specifically, we identified a connected components consisting of 40 advertisements, each published by a distinct page. As display on Figure 1, these ads advocate against the integration of Ukraine to the European Union, arguing it would create 'unfair' competition for French farmers, ads ran Jaunary 26-27th 2024, during large french farmers protest [50]. Similarly, we uncover a smaller components, made of 27 ads, displayed in Figure 7 in Appendix, published by 27 unique pages, discussing widespread corruption in Ukraine.

The group of pages advertising these ads differs from other advertisers in that they each published only a single ad within our observation window, i.e. since August 17th, 2023, and most were subsequently removed following the end of the ad campaign. Furthermore, the names of these pages follow common generation patterns, either comprising an adjective, from a limited pool, followed by 2-3 letters and a single digit (e.g., "Attraction bba2") or consisting of 4-5 words with concatenation (e.g., "Glamour GirlsHolistic HealthCommunity Connections").

Within our corpus of 320k ads published in January and February 2024, flagged as potentially political by our model, we have identified 592 unique pages matching one of this two patterns, having published 966 political ads, in 13 of the 16 countries presently considered[3]. The ads published by these identified pages predominantly target France and Germany, with 296 and 148 ads respectively. These pages have published ads present in 173 unique connected components, which collectively comprise a total of 9 491 ads.

A manual inspection of the ads published by each connected component revealed the same crypto-scam ads as previously identified and ads providing medical advice focused on incontinence. However, in France and Germany, the network of pages predominantly publishes political messages mainly related to the war in Ukraine, 80.1% and 38.5% of the ads matching the case insensitive regular expression `\bukrain\w*`.

4.2.4 pro-Russian Campaigns.

Dataset. To expand the temporal scope of our analysis, we examined the complete set of advertisements related to Ukraine, targeting France or Germany, published between August 17th, 2023 and March 31th 2024. Initially comprising 5 288 ads in France and 6 949 in Germany, matching the pattern `\bukrain\w*`, we refined this dataset by excluding non-political ads and those published by pages having launched multiple ads, as our previous analysis indicated pages involves in propaganda are of single-use.

Coordination. The dataset curated, we applied the 3Δ coordination detection method to the subset of 2 411 ads in France and 1 874 ads in Germany. We identify 393 connected components in France, encompassing a total of 2 254 ads with 580 unique texts, reaching an estimated 30.4 million user accounts between August

Figure 1: Ads, translated from French, having been ran, in January 2024, by 40 differents pages, see the detail in the Meta Ad Library.

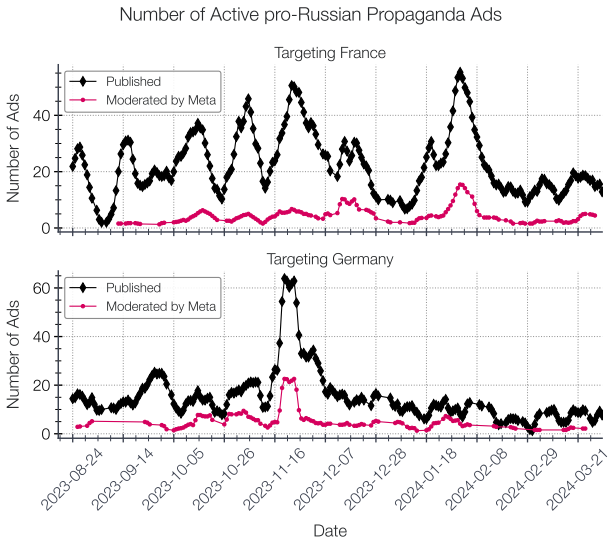

Figure 2: Number of active (black diamond markers), and moderated (red dot markers), pro-Russian propaganda ads targeting France & Germany from August 18th, 2023, to March 31th, 2024; moving average over a week.

17th, 2023, and March 31th, 2024. In Germany, we identify 283 connected components, comprising 1 572 ads with 400 unique texts, reaching approximately 7.7 million user accounts during the same period. The average reach of ads targeting France is 2.8 times larger than those targeting Germany, 13 469 reached accounts per ad vs. 4 895 accounts. On average, each connected component is made of

[3]Austria, Belgium, Bulgaria, France, Germany, Greece, Hungary, Italy, Poland, Portugal, Romania, Slovakia, Spain

ads levering 1.45 unique text variations. Withing a connected components 79.9% of the ads were published on the same day. All ads targeted only Facebook, excluding Instagram or other advertising venues.

Moderation. A manual inspection revealed that all identified 3 826 ads fall under Meta's guidelines, discussing social and geopolitical issues. Despite their political nature, none of them were self-declared as political and only 16.0% of the identified ads were moderated as political by Meta in France, and 25.5% in Germany. Moreover, as previously highlighted, moderation decisions are not consistently enforced and do not extend to duplicated ads. Furthermore, these moderation decisions are made after the ads have accumulated a total reach of 1.9 million accounts in France and 1.1 million accounts in Germany. In cases where advertisements are declared as political by the advertisers or are moderated as such by Meta, Meta discloses, in its Ad Library, a range of impressions i.e. the number of times ads having displayed on screens. Accordingly, we determine that Meta moderated ads relying pro-Russian propaganda after having displayed them between 1.7 and 2.3 million times in France and between 893,000 and 1.3 million times in Germany.

Interestingly, none of the ads contains the name of political figures; instead they refer to as "[country] leader" or use periphrasis. For instance, Joe Biden is referred to as "the American grandfather," Emmanuel Macron as "Mr 25%" (referring to his electoral results) or as "Élysée Palace" (referring to the French president's official residence), and Volodymyr Zelensky as "Ukrainian actor," "Comedian from Kiev"[4] or "the evil clown of Kiev"[5]. Moreover, we find numerous occurrences of inner-word spaces, e.g., "l ea d er." We may hypothesize these heuristics are meant to avoid detection by Meta's moderation.

Page names. The identified ads were published by page following common name generation patterns, patterns evolving over time. We display on Figure 3, the fraction of ads published daily, by Facebook page matching various name pattern. Notably, during September/October 2023, a prevalent naming convention consisted of a sequence of 6 letters followed by 'online shop' e.g. 'Ubagag online shop', observed consistently in both France and Germany. This pattern decline in prevalence by November in favour of 3 words-long name e.g. 'Classic Gardening Achievers'. In December 2023, a variety of naming patterns were observed within the network, two-words page names were predominant e.g. 'Sarah Coffee'. In January 2024, advertisements were primarily published by pages following a name pattern, previously observed by [48] in Spring 2023, 'adjective + 2-3 letters + 1 digit' e.g. 'Gorgeous lhu8' or lengthy name with concatenated words e.g. 'Movie MadnessCreative CornerFantastic Fables'. After the resurgence of two-words page names in February 2024, we observe in March 2024, a new name patterns, the concatenation of names, ending with the letter 'a' e.g. 'ElgaAkhmetova'. Overall, the naming convention used by pages promoting pro-Russia narratives evolved in sync between France and Germany.

[4]We observe two variants in French, "comique de Kiev" and "humoriste de Kiev"
[5]Used in only two ads, having reached 72,208 French accounts

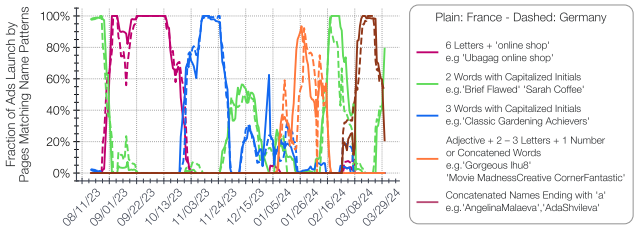

Figure 3: Fraction of ads, launch daily by Facebook pages matching a given name generation pattern, targeting France (plain line) and Germany (dashed line), moving average over a week

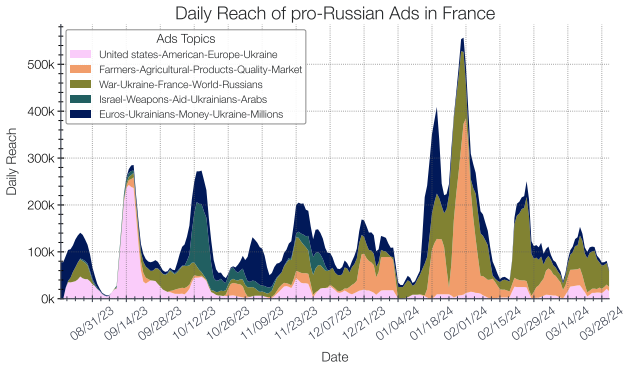

Figure 4: Daily reach, moving average over a week, of pro-Russia ads targeting France, broken down by topic, with 5 words topic descriptors.

Chronology and Narratives. Figure 2 displays the number of ads running daily from August 17th, 2023, to March 31th, 2024. We obverse a baseline, with 22.6 unique ads running daily in France and 15.0 in Germany. Performing topic modeling on the set of ads revealed five main topics for those targeting France and three for Germany. Figures 4 and 5 illustrate the daily reach of ads in France and Germany, respectively, segmented by ad topic. Daily reach is calculated as the sum of each active ad's reach, normalized by their respective campaign durations.

The reach, metric provided by Meta, may include multiple occurrences of the same account. However, it does offer insight into the motivation behind running the campaign, as reach is influenced by dedicated budgets rather than merely the number of ads created. Between August 17th, 2023, and March 31th, 2024, pro-Russian ads reached an average of 138 590 accounts daily in France and 37 326 accounts in Germany. Additionally, beyond this baseline, peaks in reach coincide with significant geopolitical events.

For instance, in the days following the announcement of a new aid package for Ukraine exceeding $1 billion by the U.S. Secretary of State on September 6th, 2023 [37], there was a notable increase in ads discussing the United States' involvement in the conflict. These ads compared the GDP of Europe and the U.S., suggesting that the U.S. instigated the conflict to bolster arms sales. Over a span of two

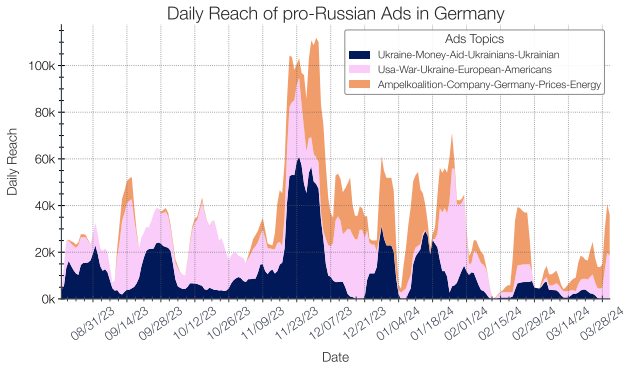

Figure 5: Daily reach, moving average over a week, of pro-Russia ads targeting Germany, broken down by topic, with 5 words topic descriptors.

days, September 8th-9th, 2023, these ads, which are detailed in the Appendix, reached 132 261 accounts in France. None of the 11 ads were moderated by Meta as political.

Promptly after the Hamas-led attack on October 7th and the subsequent resurgence of armed conflict in Gaza, a new narrative emerged accusing the Ukrainian President, referred to as the "Ukrainian comedian" of embezzling weapons from the Ukrainian army for personal gain. These weapons were alleged to have been used in attacks against Israelis. Such advertisements, detailed in the Appendix, began circulating as early as October 9th, 2023, in both France and Germany. In total, advertisements mentioning Israel garnered a combined reach of 2 134 844 in France (consisting of 222 ads with 47 unique texts) and 624 142 in Germany (comprising 149 ads with 36 unique texts). Notably, the proportion of moderated ads was higher compared to the overall ad campaign, accounting for 23.0% in France and 36.9% in Germany. However, moderation occurred after these ads had already reached a significant number of accounts, totaling 385 514 in France and 142 431 in Germany. In addition to slander the Ukrainian president, these ads advocated for disengagement of the United States in Ukraine in favor of supporting Israel. Please refer to the Appendix for examples.

On November 11th, 2023, the governing coalition led by German Chancellor Olaf Scholz agreed to double the country's military aid for Ukraine next year to 8 billion euros [49]. Subsequently, on November 21st, 2023, the German defense minister announced a €1.3 billion weapons package during a visit to Kyiv [44]. During this period, we observed a significant surge of 463 ads from the network targeting Germany. These ads, comprising 72 unique texts, criticized the German political coalition and voiced disapproval of Germany's support for Ukraine. The campaign ran from November 12th to November 30th, 2023, reaching a total of 1,676,220 accounts. Refer to Figure 13 in the Appendix for further details.

In France, we noted a peak in reach in January 2024, following the launch of the "Artillery for Ukraine" coalition between France and Ukraine [16], as well as a call from Members of the European Parliament for increased EU military aid [43]. During this time, the network focused its efforts on French farmers, advocating against the integration of Ukraine into the European Union. The ads argued that such integration would create unfair competition for French farmers and were circulated during large French farmers' protests [50]. Refer to Figure 1 for additional information.

Finally, less than two days following the attack on Crocus City Hall, advertisements relaying the Kremlin's allegations of Ukrainian and Western involvement were launched. These ads reached 41,901 French accounts between March 24th and March 31st, 2024, none were moderated by Meta; we display in Appendix the two text variations Figures 14 & 15.

To facilitate further analysis of the narratives propagated by this network, we release the full set of identified ads along with their metadata at URL.

5 DISCUSSION

Through a comprehensive analysis of the Meta Ad Library, containing *all* ads targeting EU countries, made accessible thanks to Meta's implementation of the European Union Digital Services Acts requirements, we assess how Meta enforces its political ads policy in 16 EU countries. As illustrated by the high false-positive rate in Meta's moderation [27] determining the political nature of ads is challenging and subject to disagreement between general public, advertisers and advertising platforms [54]. We started our audit with a subset of ads, unambiguously falling under Meta's guidelines on political advertising's [6], namely ads featuring the name of head of states, in Germany, France and Italy. Our analysis indicates that Meta's moderation of ads remains limited across all three countries, with substantial variations.

Subsequently, we considered the whole range of ads falling under Meta's political ads policy. In particular, we combined manual annotation with a language model trained to predict if a given advertisement text fall under Meta's guidelines on political ads. Overall, within the 16 EU countries of interest, we find that Meta's moderation is imprecise, over 60% of the ads moderated by Meta as political do not fall under its own guidelines, and that only 4.8% of undeclared political are moderated by Meta. We observe important differences in Meta's moderation recall across countries, with up to 22.8% of undeclared political ads being moderated in Ireland, the only English-speaking country in our set, and only 3.4% in Austria. Similarly, we observe difference in the fraction of political ads, detected by our model, being declared as such by their advertisers, over half of them are declared in Netherlands and Romania but only 10% in Portugal. Self-declaration of ads political nature by advertisers, lead to significant over-estimation, over the 16 countries of interest, only 53.4% of ads self-declared as political were, under Meta's guidelines, required to do so.

The findings discussed herein are aligned with previous research. Le Pochat et al. estimated Meta's moderation precision at 45% in the US in 2022, with a moderation recall of 22% computed on clearly political pages, where enforcement is likely easier as underlined in [27]. Sosnovik and Goga found that 11% of ads instinctively considered as political or strongly political by the general public were not declared as such on Meta's platform [54]. Edelson et al. observed a 79% false positive rate in Meta's moderation efforts in the US from May 2018 to June 2019 [18], highlighting challenges in accurately identifying and moderating content. Additionally, Le Pochat et al. identified significant variations in Meta's moderation

performance across different countries, with some having up to 53 times higher false negative rates among clearly political pages than in the U.S. [27]. In the context of the 2018 Brazilian general election, Silva et al. detected numerous undeclared political ads, concluding that "the main culprit for this situation is that advertisers need to self-declare their political ads as such [..]. It is not clear whether Facebook has any mechanism to enforce compliance." [53].

Furthermore, we explored among the set of undeclared political ads, as detected by our model, for coordinated advertising campaigns. To this end, we adapted a method of detection of inauthentic duplication of content developed by [52] for social media messages. Through this method, we identify set of ads, linked to each other by being a *translation, rewording or copy-pasta*. In particular, in January-February 2024, we identify, 8 232 investment-scam ads, published by 971 unique pages, targeting 10 out of 16 countries of interest. These ads, are clickbait and misleading having a common structure "The host of [TV Show] [Name of the host] described [Name of a journalist, celebrity or political figure] as irresponsible and declared live that financial information of such magnitude could shake the foundations of [country] society". Among those ads, delivered by Meta to a total of 128 millions accounts in two months, none were declared as related to social issues by their author, and less than 0.1% of them were moderated by Meta. The fraudulent nature of such ads, as well as their overall infrastructure both on Meta's product and external landing pages, has been extensively studied in [20], documenting the network in spring 2023, before our observation windows, and recently investigated in [45].

Beyond coordinated scam ads, we identify a large network of Facebook pages engaged in multi-pages coordinated campaign relaying pro-Russian propaganda. The detected network large of over 3 826 pages is active in both France and Germany, since, at least, August 17th 2023. These pages, total a reach of 30.4 millions accounts in France and 7.7 millions accounts in Germany. The average reach of ads targeting France is 2.8 times larger than those targeting Germany, reflecting the advertisers motivations, as the reach is directly associated to campaign budget. Moreover, we observe significant variation in the activity of the network, in sync with national and geopolitical events. For instance, in November 2023, while German Chancellor agreed to double the country's military aid for Ukraine next year to 8 billion euros and the German defense minister announced a €1.3 billion weapons package during a visit to Kyiv, we observe a all-time high activity of the network in Germany, reaching 1 676 220 accounts from November 12th to November 30th, 2023 with 463 ads criticizing the German political coalition and voiced disapproval of Germany's support for Ukraine. Similarly, in January 2024 in France, following the launch of the "Artillery for Ukraine" coalition between France and Ukraine [16], as well as a call from Members of the European Parliament to increase EU military aid [43]. During this time, the network focused its efforts on French farmers, advocating against the integration of Ukraine into the European Union. The ads argued that such integration would create unfair competition for French farmers and were circulated during large French farmers' protests [50].

Segments of the campaign, known as the Doppelganger operation, have previously been documented by various civil society organizations [12, 23, 26, 45, 48], the French service combating foreign digital interference [58] and by Meta itself [35]. However, by systematically examining coordinated campaigns across all undeclared political advertisements, rather than relying solely on manual, expert-based searches, a more comprehensive understanding of the network's activities can be attained. Consistent with prior findings, pages disseminating such propaganda exhibit similarities, including a common name generation pattern, evolving in sync in France and Germany. Despite its significant volume, lack of stealthiness, and previous documentation, the network's activities persist with minimal moderation from Meta, with less than 20% of their advertisements identified as political and subject to moderation.

Despite the extensive coverage of our analysis, we acknowledge several limitations. Firstly, our model was trained and used exclusively on ads from 16 EU countries, representing over 90% of ads published between August 2023 and February 2024. This approach aimed to reduce the manual annotation burden associated with the training set. However, this limited geographical focus may not fully capture the diversity of ad content and moderation practices across other EU countries.

Furthermore, although our annotators aimed to adhere closely to Meta's moderation guidelines, discrepancies emerged, as indicated by Meta's moderation high false-positive rate determined through manual annotation. These differences highlight the challenges inherent in accurately classifying political ads, especially in cases where subjective interpretation is required. Additionally, our analysis only considered the textual content of ads, while Meta's moderation system purportedly incorporates multiple modalities such as images, videos, and landing pages. Consequently, our estimates are constrained by the textual information and by the moderation guidelines publicly available to our annotators.

Future research could explore approaches that integrate multiple modalities for more robust detection of political ads. Moreover, while our focus was primarily on coordinated campaigns involving content duplication, there remains a need to investigate the entire set of undeclared political ads, for instance to characterize political actors running undeclared advertisements.

In light of the upcoming European parliamentary elections and empowered by the Digital Services Acts, the European Commission has issued guidelines for Very Large Online Platforms to address systemic online risks potentially affecting the integrity of elections [15]. These guidelines recommends political advertisements to "labelled in a clear, salient and unambiguous manner" and platforms to ensure adequate policies and systems aimed at preventing the misuse of advertising systems to disseminate misleading information, disinformation, and foreign interference in electoral processes. Our findings suggest that Meta still lack systemic and effective systems to ensure the integrity of forthcoming elections.

ACKNOWLEDGMENTS

This work was motivated by Jean Liénard talk as Fosdem'24 [30], we thank them for the fruitful discussions. Paul Bouchaud acknowledges the Jean-Pierre Aguilar fellowship from the CFM Foundation for Research. The author extent their sincere acknowledgements to LeJo for their instrumental help in the data collection, the support from AI Forensics and independent reviewers, in particular Marc Faddoul, Salvatore Romano, Sonia Tabti, Claudio Agosti.

REFERENCES

[1] Mozilla Ad Transparency. 2019. Data Collection Log — EU Ad Transparency Report. https://adtransparency.mozilla.org/eu/log/. Accessed: (01-04-2024).

[2] Transparency Center. 2024. Ads about social issues, elections or politics. https://transparency.fb.com/en-gb/policies/ad-standards/siep-advertising/siep/. Accessed: (17-03-2024).

[3] Meta Business Help Centre. [n. d.]. Reach. https://www.facebook.com/business/help/710746785663278. Accessed: (07-04-2024).

[4] Meta Business Help Centre. 2024. About ads about social issues, elections or politics. https://www.facebook.com/business/help/167836590566506?id=288762101909005. Accessed: (17-03-2024).

[5] Meta Business Help Centre. 2024. About ads in review. https://www.facebook.com/business/help/204798856225114?id=649869995454285. Accessed: (06-04-2024).

[6] Meta Business Help Centre. 2024. About social issues. https://www.facebook.com/business/help/214754279118974?id=288762101909005. Accessed: (17-03-2024).

[7] Meta Business Help Centre. 2024. Ad authorisation exemptions and how they work. https://www.facebook.com/business/help/387111852028957. Accessed: (17-03-2024).

[8] Meta Business Help Centre. 2024. Ads about social issues and selling products or services. https://www.facebook.com/business/help/287622936276216. Accessed: (17-03-2024).

[9] Meta Business Help Centre. 2024. Be authorised to run ads about social issues, elections or politics. https://www.facebook.com/business/help/208949576550051?id=288762101909005. Accessed: (17-03-2024).

[10] Meta Business Help Centre. 2024. Confirm your identity to run ads about social issues, elections or politics. https://www.facebook.com/business/help/2992964394067299?id=288762101909005. Accessed: (17-03-2024).

[11] Meta Business Help Centre. 2024. How ads about social issues, elections or politics are reviewed (with examples). https://www.facebook.com/business/help/313752069181919?id=288762101909005. Accessed: (17-03-2024).

[12] Valentin Châtelet. 2024. French prime minister faces onslaught of online attacks. https://dfrlab.org/2024/02/20/french-prime-minister-faces-onslaught-of-online-attacks/. Accessed: 29-03-2024).

[13] CLDR. 2023. Unicode Common Locale Data Repository. https://cldr.unicode.org/. Accessed: (06-04-2024).

[14] Nick Clegg. 2023. New Features and Additional Transparency Measures as the Digital Services Act Comes Into Effect. https://about.fb.com/news/2023/08/new-features-and-additional-transparency-measures-as-the-digital-services-act-comes-into-effect/. Accessed: (31-03-2024).

[15] European Commission. 2024. ECommission publishes guidelines under the DSA for the mitigation of systemic risks online for elections. https://ec.europa.eu/commission/presscorner/detail/en/ip_24_1707. Accessed: (07-04-2024).

[16] Ministère des Armées. 2024. Ukraine : la coalition artillerie est lancée (in French). https://www.defense.gouv.fr/actualites/ukraine-coalition-artillerie-est-lancee. Accessed: 29-03-2024).

[17] Matthijs Douze, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, and Hervé Jégou. 2024. The Faiss library. arXiv:2401.08281 [cs.LG]

[18] Laura Edelson, Tobias Lauinger, and Damon McCoy. 2020. A Security Analysis of the Facebook Ad Library. In *2020 IEEE Symposium on Security and Privacy (SP)*. 661–678. https://doi.org/10.1109/SP40000.2020.00084

[19] Emilio Ferrara. 2023. Social bot detection in the age of ChatGPT: Challenges and opportunities. *First Monday* (June 2023). https://doi.org/10.5210/fm.v28i6.13185

[20] Check First. 2024. Facebook Hustles: The Hidden Mechanics of a Scam Machinery Impersonating News Organisations and Creators. https://checkfirst.network/facebook-hustles-the-hidden-mechanics-of-a-scam-machinery-impersonating-news-organisations-and-creators/. Accessed: (17-03-2024).

[21] French Ambassador for Digital Affairs. 2019. Facebook Ads Library Assessment. https://disinfo.quaidorsay.fr/en/facebook-ads-library-assessment. Accessed: (01-04-2024).

[22] Phillipa Gill, Vijay Erramilli, Augustin Chaintreau, Balachander Krishnamurthy, Konstantina Papagiannaki, and Pablo Rodriguez. 2013. Best paper – Follow the money: understanding economics of online aggregation and advertising. In *Proceedings of the 2013 conference on Internet measurement conference (IMC'13)*. ACM. https://doi.org/10.1145/2504730.2504768

[23] ISD. 2022. Pro-Kremlin Network Impersonates Legitimate Websites and Floods Social Media with Lies. https://www.isdglobal.org/digital_dispatches/pro-kremlin-network-impersonates-legitimate-websites-and-floods-social-media-with-lies/. Accessed: (01-04-2024).

[24] Armand Joulin, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Hérve Jégou, and Tomas Mikolov. 2016. FastText.zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651* (2016).

[25] Levi Kaplan, Nicole Gerzon, Alan Mislove, and Piotr Sapiezynski. 2022. Measurement and analysis of implied identity in ad delivery optimization. In *Proceedings of the 22nd ACM Internet Measurement Conference (IMC '22)*. ACM. https://doi.org/10.1145/3517745.3561450

[26] EU Disinfo Lab. 2022. Doppelganger – Media clones serving Russian propaganda. https://www.disinfo.eu/doppelganger/. Accessed: (01-04-2024).

[27] Victor Le Pochat, Laura Edelson, Tom Van Goethem, Wouter Joosen, Damon McCoy, and Tobias Lauinger. 2022. An Audit of Facebook's Political Ad Policy Enforcement. In *Proceedings of the 31st USENIX Security Symposium, Security 2022 (Proceedings of the 31st USENIX Security Symposium, Security 2022)*. USENIX Association, 607–624. Publisher Copyright: © USENIX Security Symposium, Security 2022.All rights reserved.; 31st USENIX Security Symposium, Security 2022 ; Conference date: 10-08-2022 Through 12-08-2022.

[28] Daniel D. Lee and H. Sebastian Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 6755 (oct 1999), 788–791. https://doi.org/10.1038/44565

[29] LeJo. 2024. Copy of the ads from Facebook Ad Library API. https://github.com/Lejo1/facebook_ad_library. Accessed: (02-03-2024).

[30] Jean Liénard. 2024. Detecting Propaganda on Facebook and Instagram Ads using Meta API. Talk at FOSDEM'24, Brussels.

[31] Gurmeet Singh Manku, Arvind Jain, and Anish Das Sarma. 2007. Detecting near-duplicates for web crawling. In *WWW 2007 (16th International Conference on the World Wide Web)*. Banff, 141–150. http://doi.acm.org/10.1145/1242572.1242592

[32] Meta. 2018. A New Level of Transparency for Ads and Pages. https://about.fb.com/news/2018/06/transparency-for-ads-and-pages/. Accessed: (31-03-2024).

[33] Meta. 2018. Shining a Light on Ads With Political Content. https://about.fb.com/news/2018/05/ads-with-political-content/. Accessed: (31-03-2024).

[34] Meta. 2019. Protecting Elections in the EU. https://about.fb.com/news/2019/03/ads-transparency-in-the-eu/. Accessed: (01-04-2024).

[35] Meta. 2023. Meta Adversarial Threat Report - Q2. https://transparency.fb.com/sr/Q2-2023-Adversarial-threat-report. Accessed: (07-04-2024).

[36] Meta. 2024. Meta Ad Library API. https://m.facebook.com/ads/library/api/. Accessed: (17-03-2024).

[37] US Departement of State. 2023. Secretary Blinken's Travel to Ukraine. https://www.state.gov/secretary-blinkens-travel-to-ukraine-2/. Accessed: 29-03-2024).

[38] Council of the EU. 2024. EU introduces new rules on transparency and targeting of political advertising. https://www.consilium.europa.eu/en/press/press-releases/2024/03/11/eu-introduces-new-rules-on-transparency-and-targeting-of-political-advertising/. Accessed: (02-04-2024).

[39] OpenAI. 2023. OpenAI Models. https://platform.openai.com/docs/models/gpt-3-5-turbo. Accessed: (17-03-2024).

[40] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. arXiv:2203.02155 [cs.CL]

[41] Derek O'Callaghan, Derek Greene, Joe Carthy, and Pádraig Cunningham. 2015. An analysis of the coherence of descriptors in topic modeling. *Expert Systems with Applications* 42, 13 (Aug. 2015), 5645–5657. https://doi.org/10.1016/j.eswa.2015.02.055

[42] Panagiotis Papadopoulos, Nicolas Kourtellis, Pablo Rodriguez Rodriguez, and Nikolaos Laoutaris. 2017. If you are not paying for it, you are the product: how much do advertisers pay to reach you?. In *Proceedings of the 2017 Internet Measurement Conference (IMC '17)*. ACM. https://doi.org/10.1145/3131365.3131397

[43] European Parliament. 2024. Ukraine war: MEPs to call for more EU military aid. https://www.europarl.europa.eu/news/en/agenda/briefing/2024-02-26/6/ukraine-war-meps-to-call-for-more-eu-military-aid. Accessed: 29-03-2024).

[44] Politico. 2023. German defense minister announces €1.3B weapons package during visit to Kyiv. https://www.politico.eu/article/german-defense-minister-announces-new-military-aid-package-during-his-visit-to-kyiv/. Accessed: 30-03-2024).

[45] Qurium. 2024. A Journey into the Crypt of Cloned Media. https://www.qurium.org/alerts/into-the-crypt-of-cloned-media/. Accessed: 29-03-2024).

[46] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. http://arxiv.org/abs/1908.10084

[47] Nils Reimers and Iryna Gurevych. 2020. Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. arXiv:2004.09813 [cs.CL]

[48] Reset. 2023. Vast Networks of Fake Accounts Raise Questions About Meta's Compliance with the EU's DSA. https://www.reset.tech/resources/vast-networks-of-fake-accounts-raise-questions-about-meta-s-compliance-with-the-eu-s-new-digital-rulebook/. Accessed: (17-03-2024).

[49] Reuters. 2023. Germany set to double Ukraine military aid. https://www.reuters.com/world/europe/germany-set-double-its-ukraine-military-aid-under-scholz-plan-bloomberg-news-2023-11-12/. Accessed: 30-03-2024).

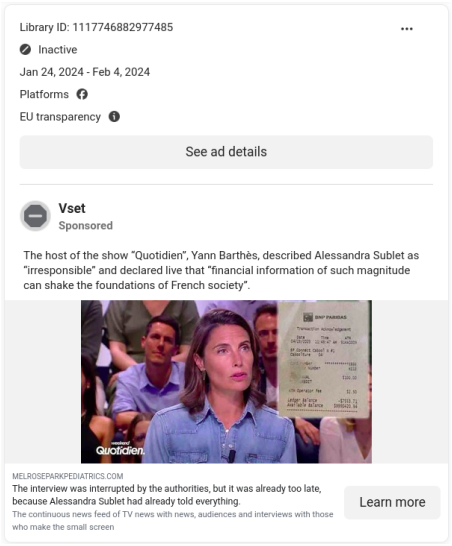

Figure 6: Ads initially run in France, text has been automatically translated to English, see the detail in the Meta Ad Library

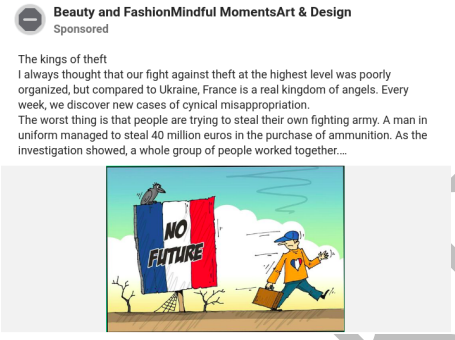

Figure 7: Ads initially run in France, text has been automatically translated to English, see the detail in the Meta Ad Library

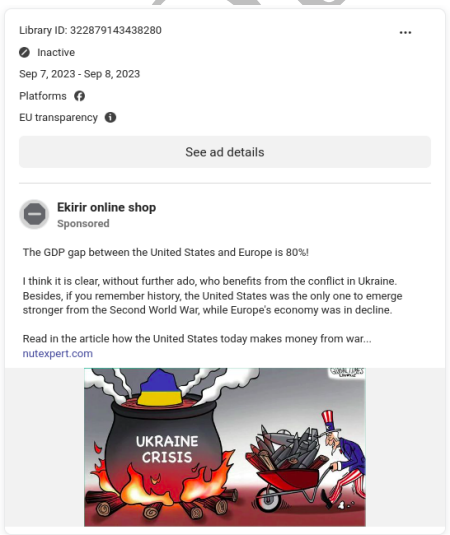

Figure 8: Ads initially run in France, text has been automatically translated to English, see the detail in the Meta Ad Library

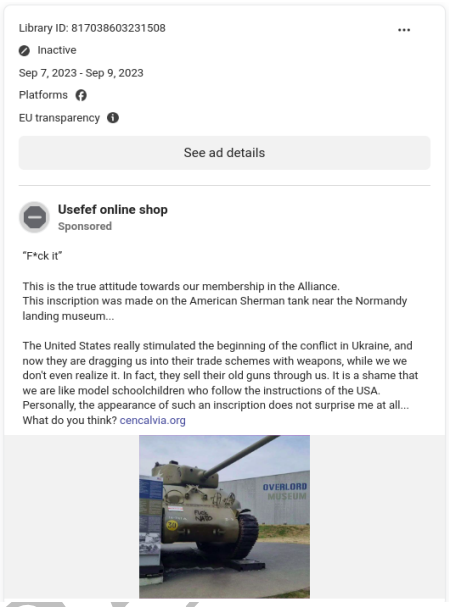

Figure 9: Ads initially run in France, text has been automatically translated to English, see the detail in the Meta Ad Library

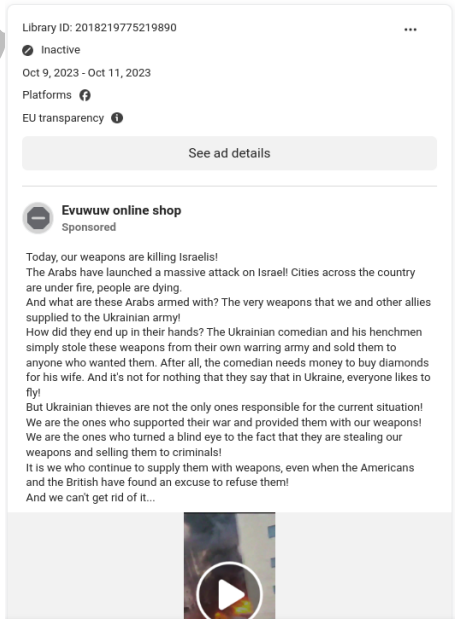

Figure 10: Ads initially run in France, text has been automatically translated to English, see the detail in the Meta Ad Library

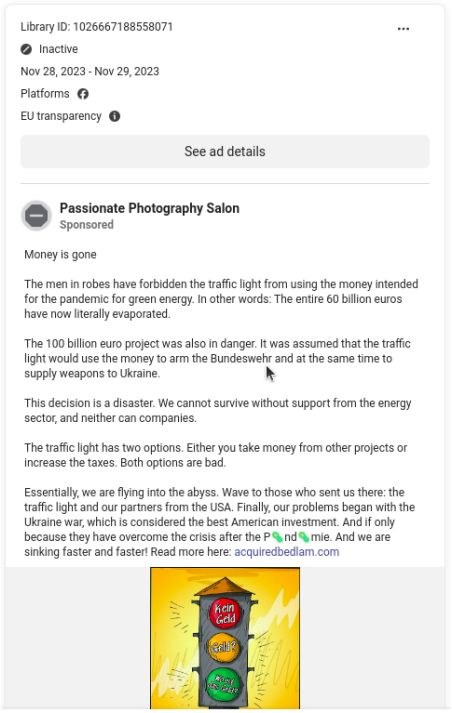

Figure 13: Ads initially run in Germany, text has been automatically translated to English, see the detail in the Meta Ad Library

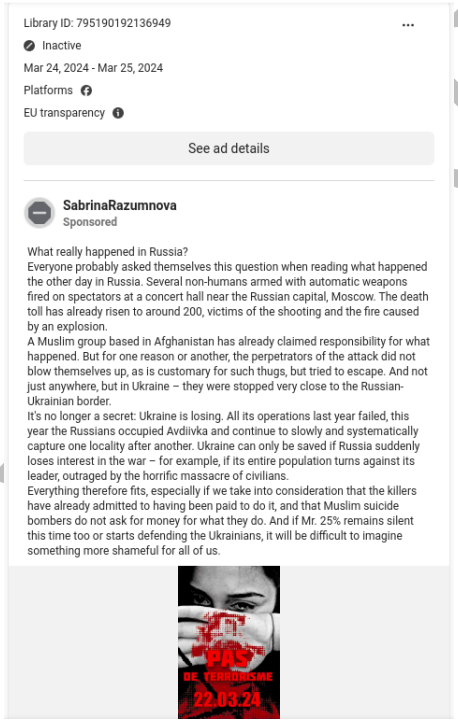

Figure 14: Ads initially run in France, text has been automatically translated to English, see the detail in the Meta Ad Library

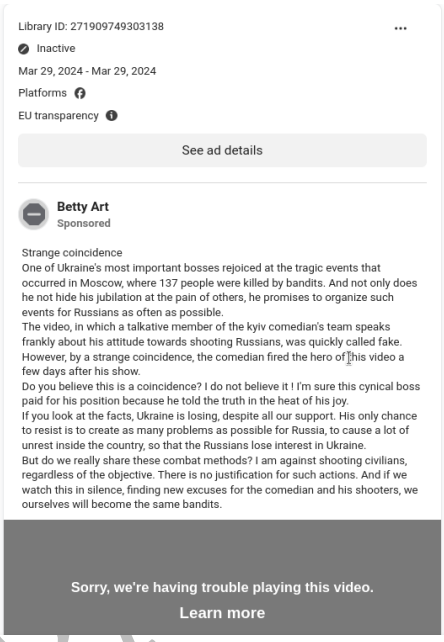

Figure 15: Ads initially run in France, text has been automatically translated to English, see the detail in the Meta Ad Library

[50] Reuters. 2024. French farmers block roads, dump produce as protest edges closer to Paris. https://www.reuters.com/world/europe/french-farmers-damage-overseas-goods-protests-continue-2024-01-25/. Accessed: 30-03-2024).
[51] Filipe N. Ribeiro, Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnnatan Messias, Fabricio Benevenuto, Oana Goga, Krishna P. Gummadi, and Elissa M. Redmiles. 2019. 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 (FAT* '19)*. ACM. https://doi.org/10.1145/3287560.3287580
[52] Manon Richard, Lisa Giordani, Cristian Brokate, and Jean Liénard. 2023. Unmasking information manipulation: A quantitative approach to detecting Copy-pasta, Rewording, and Translation on Social Media. arXiv:2312.17338 [cs.SI]
[53] Márcio Silva, Lucas Santos de Oliveira, Athanasios Andreou, Pedro Olmo Vaz de Melo, Oana Goga, and Fabricio Benevenuto. 2020. Facebook Ads Monitor: An Independent Auditing System for Political Ads on Facebook. In *Proceedings of The Web Conference 2020 (WWW '20)*. ACM. https://doi.org/10.1145/3366423.3380109
[54] Vera Sosnovik and Oana Goga. 2021. Understanding the Complexity of Detecting Political Ads. In *Proceedings of the Web Conference 2021 (WWW '21)*. ACM. https://doi.org/10.1145/3442381.3450049
[55] Till Speicher, Muhammad Ali, Giridhari Venkatadri, Filipe Nunes Ribeiro, George Arvanitakis, Fabrício Benevenuto, Krishna P. Gummadi, Patrick Loiseau, and Alan Mislove. 2018. Potential for Discrimination in Online Targeted Advertising. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research, Vol. 81)*, Sorelle A. Friedler and Christo Wilson (Eds.). PMLR, 5–19. https://proceedings.mlr.press/v81/speicher18a.html
[56] The New York Times. 2019. Ad Tool Facebook Built to Fight Disinformation Doesn't Work as Advertised. https://www.nytimes.com/2019/07/25/technology/facebook-ad-library.html. Accessed: (01-04-2024).
[57] Lewis Tunstall, Nils Reimers, Unso Eun Seo Jo, Luke Bates, Daniel Korat, Moshe Wasserblat, and Oren Pereg. 2022. Efficient few-shot learning without prompts. *arXiv preprint arXiv:2209.11055* (2022).
[58] Viginum. 2023. A complex and persistent information manipulation campaign. https://www.sgdsn.gouv.fr/publications/maj-19062023-rrn-une-campagne-numerique-de-manipulation-de-linformation-complexe-et. Accessed: (01-04-2024).

[59] Zihan Zhang, Meng Fang, Ling Chen, and Mohammad Reza Namazi Rad. 2022. Is Neural Topic Modelling Better than Clustering? An Empirical Study on Clustering with Contextual Embeddings for Topics. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.naacl-main.285

Country	# Ads	# Pages	# Unique Ads	Total Reach
Austria	69	3	2	13995
Belgium	75	8	4	302557
France	2658	391	36	61486015
Germany	65	8	4	166471
Greece	70	11	4	1303348
Hungary	13	1	1	102185
Italy	2836	288	52	36153205
Netherlands	115	6	4	103659
Portugal	249	56	6	4023988
Spain	2082	241	45	24533229

Table 3: Investment Scam Ads Statistics by Country.

Received 9 April 2024

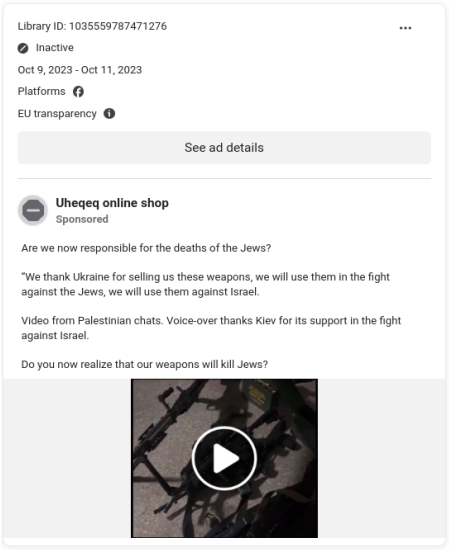

Figure 11: Ads initially run in Germany, text has been automatically translated to English, see the detail in the Meta Ad Library

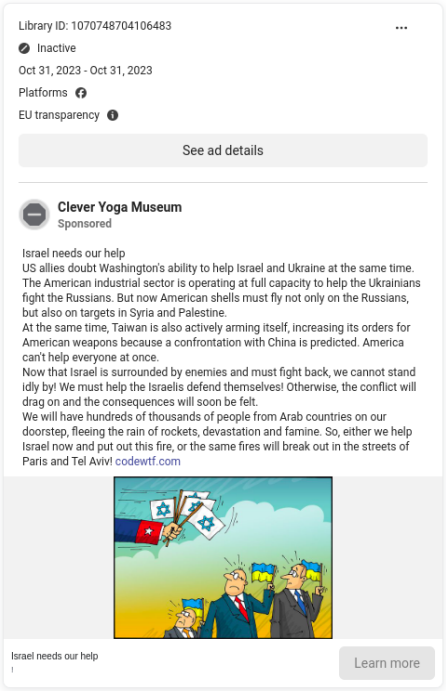

Figure 12: Ads initially run in France, text has been automatically translated to English, see the detail in the Meta Ad Library