

# Detecting Propaganda on Facebook and Instagram Ads using Meta API

VIGINUM - Jean Liénard

---

# Viginum

A technical and operational government department against foreign digital interference



Detecting and analyzing  
*foreign digital  
interference*



Leading and coordinating  
*the French inter-ministerial  
protection mechanism*



Contributing  
*to European and  
international work*

# Why do we care about Ads

Micro-targeting scandals



Cambridge  
Analytica

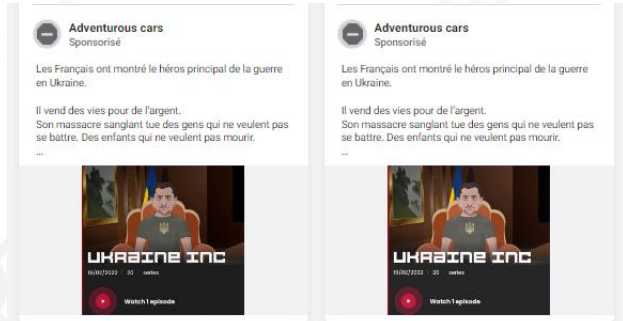
2016 Trump campaign

Brexit

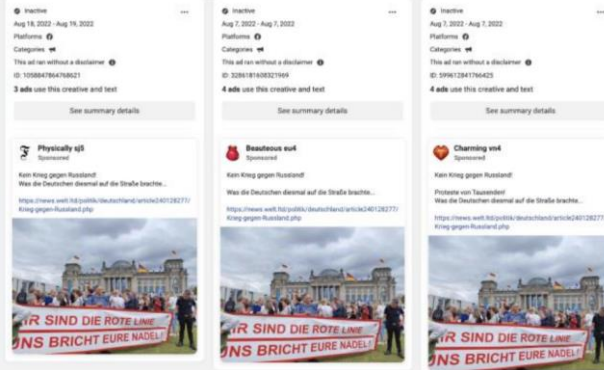
Others?

# Why do we care about Ads

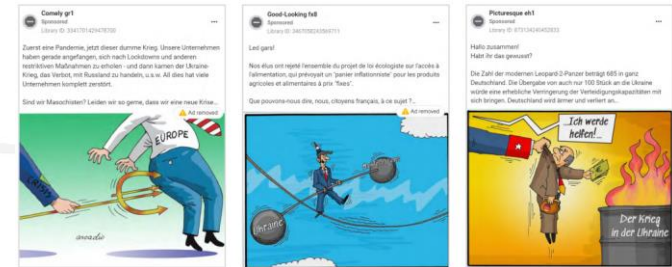
Long-running disinformation campaigns such as Doppelganger



June 2023 : Viginum report



Sept 2022 : EU Disinfo Lab report



Oct 2023 : Reset report

# Why do we care about Ads

## Blue compass 15 Sponsorisé

Une fois de plus, la France devra faire des sacrifices  
Les agriculteurs exigent le renforcement des règles  
d'importation des produits agricoles en provenance  
d'Ukraine. Les produits ukrainiens sont très bon marché  
et nos producteurs ne peuvent pas résister à la  
concurrence.  
Par ailleurs, les produits ukrainiens sont de très...



Une fois de plus, la France devra faire des sacrifices  
However, ducks have begun to rent chickens over  
the past few months, specifically for...

En savoir plus

## Classic Travel Dreamers Sponsorisé

Une fois de plus, la France devra faire des sacrifices  
Les agriculteurs exigent le renforcement des règles  
d'importation des produits agricoles en provenance  
d'Ukraine. Les produits ukrainiens sont très bon marché  
et nos producteurs ne peuvent pas résister à la  
concurrence.  
Par ailleurs, les produits ukrainiens sont de très...



Une fois de plus, la France devra faire des sacrifices  
However, ducks have begun to rent chickens over  
the past few months, specifically for...

En savoir plus

Last week

# How « open » are Facebook Ads?

Two parts:

- I) **A practical guide** on how to access the data
- II) **A study of available data** across different categories

# Practical information

## 3+1 ways to access to ads

### 1. Meta Daily report

- Download .csv file, but limited to political advertising

### 2. Meta Web portal

- Browse everything, but finicky and slow

### 3. Meta API

- Access everything, but requires registration and coding

### 4. (bonus) open-source mirror

- Code on [https://github.com/Lejo1/facebook\\_ad\\_library](https://github.com/Lejo1/facebook_ad_library) + Data Mirror hosted on Kaggle

#### Download a full report

Customize the report (.csv) by choosing the date range and end date (within the most recently available 7 days).

Time range

Last day ▾

End date

📅 Jan 21, 2024

Jan 21, 2024

Download report

#### Ad Library

Search all the ads currently running across Meta technologies, as well as:

- Ads about social issues, elections or politics that have run in the past seven years
- Ads that have run anywhere in the EU in the past year

To find an ad, search for keywords or an advertiser.



#### Search ads

Set your location and choose an ad category to start your search. [View search tips.](#)

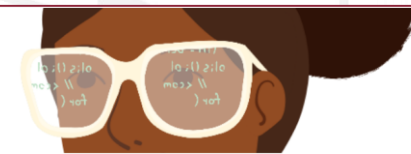
France ▾

Ad category ▾

🔍 Choose an ad category

Access the API

[Release notes](#)



Meta Ad Library API

# Howto API [1/2]

## Registration process

Two steps :

1. Verify Meta Developers account
  - provide a phone number
  - OR provide a credit/debit card info
2. Verify your official identity
  - Provide your official ID document (government's ID card, passport, ...)



# Howto API [2/2]

## (un)documented features

- Retrieve all ads =
  - `search_term='\"*\"'`
- Per-token request quota
  - but no limit on number of tokens
  - ⇒ rotate over a collection of `<ACCESS_TOKEN>`
- Limit on returned ads
  - By default `limit=20`, but you can set up to `limit=5000`
  - Depending on the payload size, this will fail: you have to retry with a lower `limit`

CURL

```
curl -G \  
-d "search_terms='california'" \  
-d "ad_type=POLITICAL_AND_ISSUE_ADS" \  
-d "ad_reached_countries=['US']" \  
-d "access_token=<ACCESS_TOKEN>" \  
"https://graph.facebook.com/<VERSION>/ads_archive"
```

# Available Data

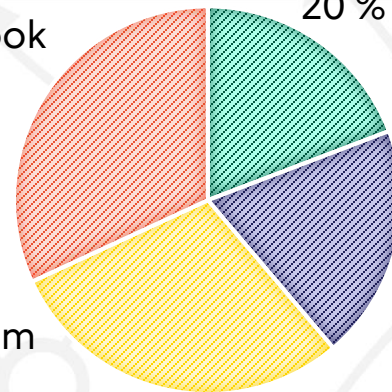
Scope: French language ads targeting France

Facebook  
32 %

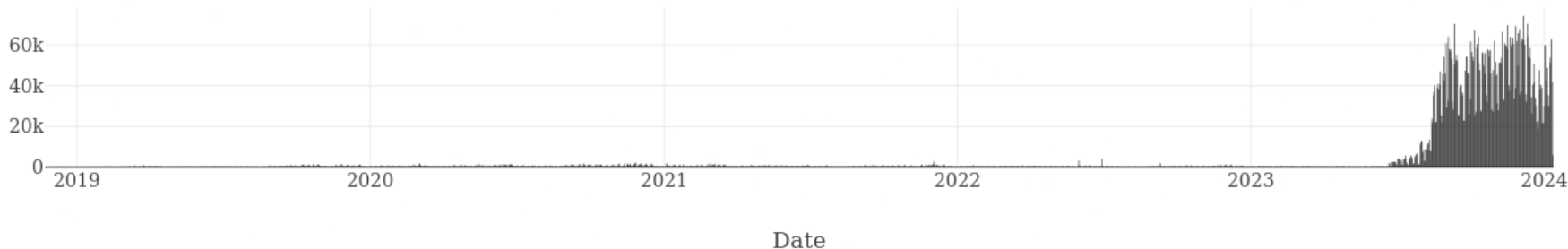
Audience Network  
20 %

Messenger  
19 %

Instagram  
29 %



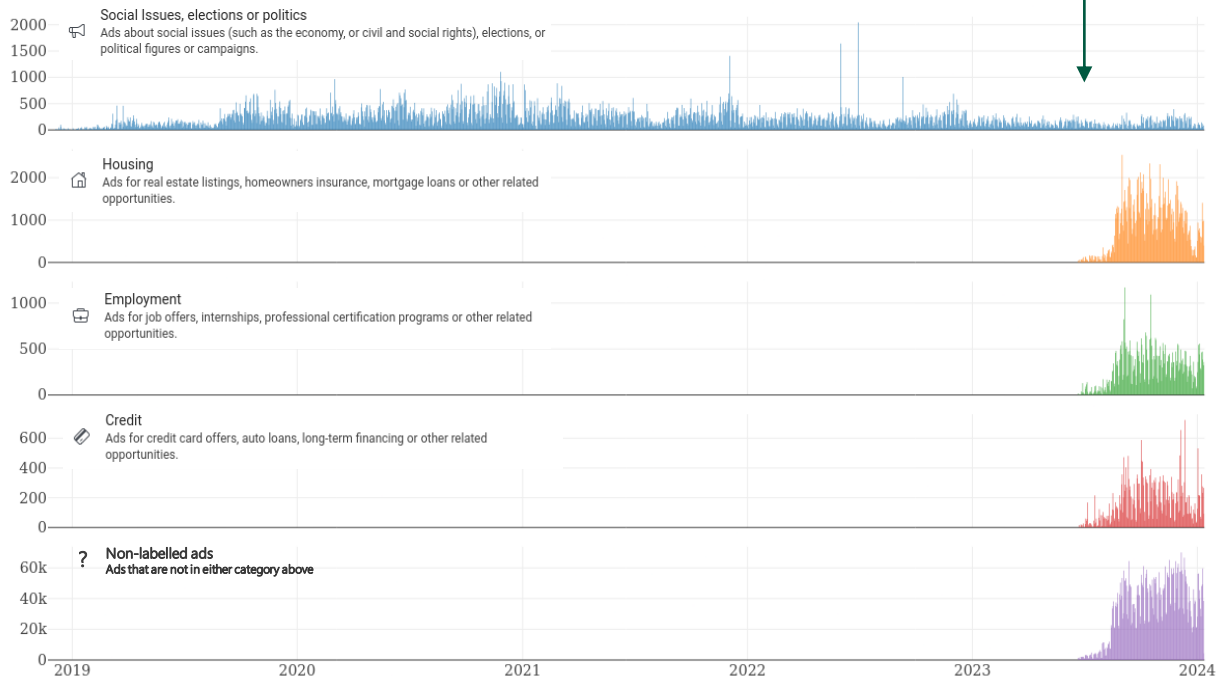
Available ads through time



# Available Data

## Breakdown:

Available ads through time



2023/08/17:

A change in the API makes non-political ads available

~ 0.3 % of recent ads

2.3 %

0.8 %

0.5 %

96.1 %

# Special case of political ads

## A manipulation vector

*Social Issues, elections or politics* is a **self-declared category**.

Ads in this category:

- are from a real person (it requires sending official ID)
- have more information available (currency/amount paid, specific demographics targeted)
- may be subject to additional scrutiny from Facebook

⇒ *Foreign information manipulations do not check the box*

### Categories

Select the categories that best describe what this campaign will advertise.

Declare category if applicable



#### Credit

Ads for credit card offers, auto loans, long-term financing or other related opportunities.



#### Employment

Ads for job offers, internships, professional certification programs or other related opportunities.



#### Housing

Ads for real estate listings, homeowners insurance, mortgage loans or other related opportunities.



#### Social Issues, elections or politics

Ads about social issues (such as the economy, or civil and social rights), elections, or political figures or campaigns.

[Get help declaring categories](#)

# Special case of political ads

Can we infer what label should have been used?

Two goals :

1. Identify automatically Doppelgänger ads
2. Estimate how many French ads should really be labeled as *Social Issues, Elections or Politics*

*Past estimates exist in the scientific litterature in other countries:*

- 2.2% mislabeling in Brazil (Silva et al. 2020)
- From 2 to 4% mislabeling in the US (Sosnovik and Goga, 2021)

**Let's build a model!**

# Special case of political ads

Can we infer what label should have been used?

## On Detecting Policy-Related Political Ads: An Exploratory Analysis of Meta Ads in 2022 French Election

Vera Sosnovik<sup>\*</sup>  
Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG  
France

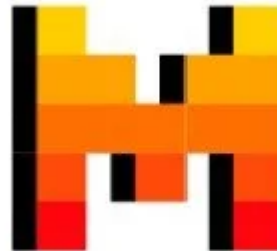
Maximin Coavoux  
Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG  
France

Romaissa Kessi<sup>\*</sup>  
Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG  
France

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

(WWW '23), May 1–5, 2023, Austin, TX, USA. ACM, New York, NY, USA,  
12 pages. <https://doi.org/10.1145/3543507.3583875>

+



Academic dataset containing  
manually annotated French-language  
political ads (Sosnovik et al. 2023)

Mistral-7B LLM with the *unsloth* python library  
(1h training on a free-to-use T4 Google Colab)

# Special case of political ads

Can we infer what label should have been used?

- Process:

## 4016 political ads:

- From supplemental material of Sosnovik et al. 2023
- Collected in early 2022 from user browsing sessions and classified into 9 fine-grained categories

## 4016 control ads:

- Collected with the API using *country=FR* and *language=fr*
- Random sample in late 2023, picking only non-labeled ads

**Split:**  
90% train  
10% test

## Fine-tuning of Mistral-7B to classify in 10 categories:

- one of the 9 political categories
- or « non political »

- Outcome:

- **91% precision in forecasting « should have been labeled as political »**

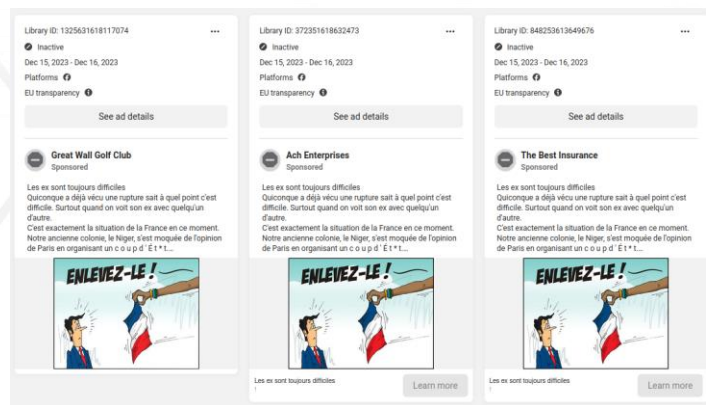
# Special case of political ads

Can we infer what label should have been used?

Application to Doppelganger:

- 100% of ads are detected as mislabeled
  - 88% as « *International affairs* »
  - 5% as « *Governmental operations* »
  - 5% as « *Energy* »
  - 2% as both « *International affairs* » and « *Economic* »

⇒ The model is able to identify mislabeled ads in Doppelganger



SYSTEM: You are a French human annotator specialized in analyzing and classifying the domain of ads. You label the type of facebook ads by putting one or more digits with the following coding: 1="International affairs", 2="Energy", 3="Governmental operations", 4="Cultural policy", 5="Social policy", 6="Healthcare", 7="Human rights, public freedom and discriminations", 8="Environnement", 9="Economic", 0="Others".

ADVERTISEMENT: Les ex sont toujours difficiles  
Quiconque a déjà vécu une rupture sait à quel point c'est difficile. Surtout quand on voit son ex avec quelqu'un d'autre.  
C'est exactement la situation de la France en ce moment. Notre ancienne colonie, le Niger, s'est moquée de l'opinion de Paris en organisant un coup d'État.  
Imaginez, ils sont tellement audacieux qu'ils ne nous demandent même pas qui sera leur chef. Ils décident tout seuls maintenant ! Ils ont également mis la mission de l'UE à la porte comme un vilain chien ! Mais même cela ne leur a pas suffi ! Ils ont également accueilli solennellement une délégation de... la Russie. Cette même Russie contre laquelle nous sommes en conflit depuis plus d'un an déjà ! La France n'a pas connu une telle humiliation depuis 1940, lorsque les Allemands sont entrés dans Paris. Je suis sûr qu'il y aura encore beaucoup d'autres humiliations. Après tout, nous n'avons pas appris à négocier avec l'Afrique comme le font les Russes. [punpirom.com](http://punpirom.com)

LIST OF DOMAINS: [1]



# Special case of political ads

Can we infer what label should have been used?

Application to random non-labeled ads:

- At least 1.9 % of non-labeled ads should be labeled as « *Social Issues, Elections or Politics* »

How can we infer this?

1. Run the model on a random sample of non-labeled ads
2. Double-check the ads identified by the model with 2 human annotators trained on Meta guidelines

- Cross-link with the proportion of labeled ads (~0.3 %) and non-labeled ads (~ 96.1 %)

⇒ Only 15 % of Social Issues, Elections and Politics ads are well labeled



Well-labeled  
political ads :  
15 %

Non-labeled,  
misleading ads :  
85 %

# Conclusion: How open are Facebook Ads?

- **Data access is non-trivial but doable**

*It is broadly relevant for the civil society!*

*Roughly 6 months of fully open data is available*

*An open-source mirror makes it easier to retrieve and analyze*

- **Self-labeling of « social issues, elections or politics » is highly problematic**

*We estimate that only 15 % of « social issues, elections or politics » ads are correctly labelled*

*⇒ 85 % of the ads escape the enhanced scrutiny associated with the label*