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

- 2016 Trump campaign
- Brexit
- Others?
Why do we care about Ads

Long-running disinformation campaigns such as Doppleganger

Sept 2022 : EU Disinfo Lab report

Oct 2023 : Reset report

June 2023 : Viginum report

Sept 2022 : EU Disinfo Lab report

Oct 2023 : Reset report

June 2023 : Viginum report
Why do we care about Ads

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 + Data Mirror hosted on Kaggle
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`
Available Data
Scope: French language ads targeting France

- Facebook: 32%
- Instagram: 29%
- Audience Network: 20%
- Messenger: 19%

Available ads through time

Date

0 20k 40k 60k
2019 2020 2021 2022 2023 2024
Available Data

Breakdown:

- Non-labelle Ads: Ads that are not in either category above

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%

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

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

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

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

Non-labelled ads
- Ads that are not in either category above

Date
- 2019
- 2020
- 2021
- 2022
- 2023
- 2024
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*
Special case of political ads

Can we infer what label should have been used?

Two goals:

1. Identify automatically Doppleganger 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?

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 Doppleganger:

- 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 Doppleganger
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
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