

Accessing platform data for research: sharing some experiences from the London Social Media Observatory



Andreu Casas

Royal Holloway Univ. of London
London Social Media Observatory



An example: studying the **intimidation of politicians** online

The
Electoral
Commission

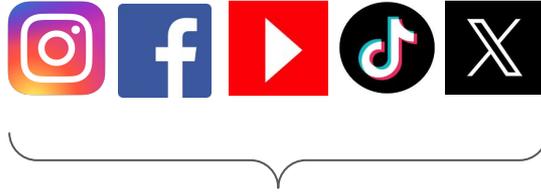


LSMO

LONDON
SOCIAL MEDIA
OBSERVATORY

- **How much** toxic speech directed at politicians online?
- **Who** is more likely to be the target: ideology, gender, etc.?
- Are some posts from politicians **inciting** hate?
- What **effect** does it have on politicians?
- How do **platforms moderate** toxic speech directed at politicians?

Collecting **comments** to **posts** from UK MPs in 5 main platforms



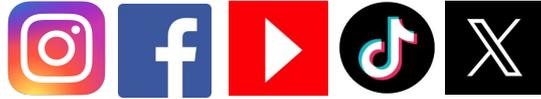
Posts (incl. visuals) from UK MPs (*Several times a day*)

All **comments** to any new post (*Every ~30 min., for ~7 days*)

Check if **previous comments** still up

Annotate some posts and comments
Fine-tune LLMs to identify toxic content

Data collection: **Research APIs** + **Scraping**



Data collection: **Research APIs + Scraping**



Dear Applicant,

Thank you for submitting an application for Researcher Access to the X API. We note that this application is for a narrow set of Researchers defined by Article 40(12) of the EU Digital Services Act.

We have reviewed your application, including the additional information you provided, and it has been denied for the following reasons:

It does not appear your organization is affiliated with a research organization within the EU, as required by Art. 40(8) of the Digital Services Act.

Additionally, your application failed to demonstrate that the access to the data requested is necessary for, and proportionate to, the purposes of your research. Data requested and scope of potential research topics was overly broad and not sufficiently tailored to defined research projects pertaining to systemic risks under Art. 34(1).

We appreciate your interest in the X Developer Platform. For general academic research or API access for other purposes, please review the different API tiers available at developer.x.com.

Thank you,

X Developer Team

Data collection: **Research APIs + Scraping**



- **Meta Content Library** (API & UI)
- **Access** took forever!
- Several **legal documents** had to be signed by University: took forever again
- Only access to posts and comments from **public** pages/profiles/groups
- Scraping/analysis of **visual data** not possible/feasible
- API **secure environment** totally useless if data/models can't be moved in/out
- Now requesting access to data from **private** users: via **DSA Data Access Portal**

Data collection: **Research APIs + Scraping**



- **YouTube Research API** (80,000 calls/day → v. 10,000 calls/day regular DevAPI)
- **Sufficient** for this project, but not for others, where we have 1+ key
- Relatively easy/fast **process**
- **Video scraping** for capturing visual content, but also to get transcripts
- ... it is getting more and more difficult

Data collection: **Research APIs + Scraping**



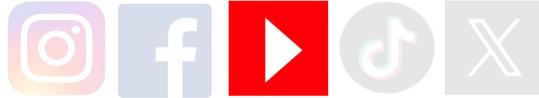
- **TikTok Research API**
- **Quick** access process (caveat: need IRB letter): only 3 days for our last request
- 1,000 API calls/day: **sufficient** for this project, but not for others
- They are OK with **many requests/keys** from several team members (reusing docs)
- **Very responsive**: we've had many calls to talk about **API issues**
- Main limitation: new **posts** take 24-48 hours to be accessible (not for **comments**)
- **Scraping video** content works fairly well
- What will happen once **US operations** taken over by new company?

Some **very** preliminary **results**



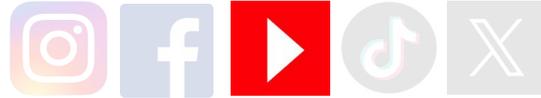
- 180 MPs → 1,105 videos → 16,557 comments (1 month: Nov. 2025)
- Quick analysis of comment toxicity using **Detoxify** model

Some **very** preliminary **results**



- 180 MPs → 1,105 videos → 16,557 comments (1 month: Nov. 2025)
- Quick analysis of comment toxicity using **Detoxify** model
 - 20.1%** of comments classified as **toxic**

Some **very** preliminary **results**

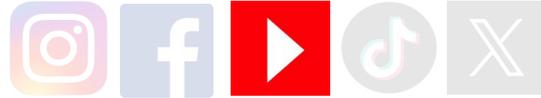


- 180 MPs → 1,105 videos → 16,557 comments (1 month: Nov. 2025)
- Quick analysis of comment toxicity using **Detoxify** model

Videos with the
most toxic comments

video_title	n_comments	n_toxic	channel_name
Parliament just voted against my Bill to leave...	2997	718	Nigel Farage
We Need To Talk About Tower Hamlets. It's Time...	2477	340	Nigel Farage
I hope every far left protester watches this v...	1070	279	Nigel Farage
Ed Miliband FAILS to turn up to vote to CUT el...	1182	193	Claire Coutinho
Hard truths to a large group of left-wing MPs ...	1049	170	Rupert Lowe MP
My economic vision for Britain.	1098	102	Nigel Farage
LIVE: Nigel Farage offers a solution to the fa...	542	79	Nigel Farage
This report commissioned and endorsed by the M...	244	52	Chris Philp MP
This is living proof that the BBC is infected ...	352	44	Nigel Farage
Keir Starmer explains what makes him proud to ...	133	41	Keir Starmer

Some **very** preliminary **results**



- 180 MPs → 1,105 videos → 16,557 comments (1 month: Nov. 2025)
- Quick analysis of comment toxicity using **Detoxify** model

text	example_type
Fucking traitor. Building fighter jets for terrorist.	toxicity_high
We the English Public ARE AWAKE, ITS TIME FOR YOU LOT TO WAKE THE FUCK UP AND GET THEM OUT!!!! ISLAM OUT !!!!!!!	severe_toxicity_high
Fucking tell em thats exactly why they do it. Bomb these countries already	obscene_high
ECONOMY IS FUCKED YOU LYING CHARLATAN	insult_high
I would happily punch he's face off..	threat_high
Every muslim should be deported	identity_attack_high
This man is a complete joke after all the mess he has caused in the UK he has the cheek to talk about climate control he is a complete w@nker đŸ˜,	toxicity_near_threshold
Greed why pay for the immigrants to stay here that saves 5 million a day get shit and leave the tax they should be in jail this labour government fucking criminals	severe_toxicity_near_threshold
Always said plaid cymru was the other Labour if you vote them in Wales all bastard's liers makes me sick to my stomach.đŸ˜-đŸ˜	obscene_near_threshold
How the fk is stop and search racist if they stop everyone ? What is fkn racist is the Black and Muslim c*nts who are killing only white people !	insult_near_threshold
I swear to my god I will personally go after you lefties as you are complicit in bringing this country to his knees.knock knock.	threat_near_threshold
pathetic, why are there even muslim mp's , stop this nonsense of putting up with islam	identity_attack_near_threshold

Accessing platform data for research: sharing some experiences from the London Social Media Observatory



Thanks!

Andreu Casas

Royal Holloway Univ. of London
London Social Media Observatory

