STUDYING SOCIAL MEDIA BEYOND THE API: USING CLONE SOCIAL MEDIA ENVIRONMENTS

Dr Aybuke Atalay

17.11.2026



AGENDA

1) Studying social media after the API changes

2) What is a clone social media environment?

3) Setting up a clone social media environment: walkthrough and a live demo

4) What can we measure and manipulate? Results from an example study

5) Pushback and limitations

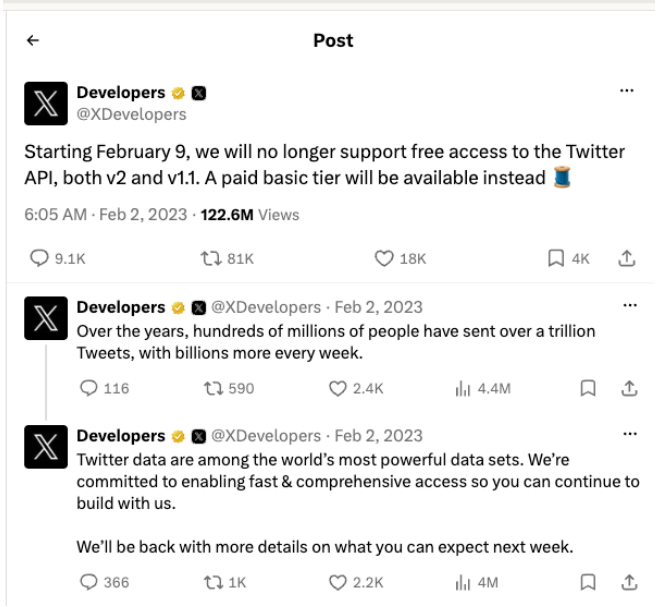
Part 1

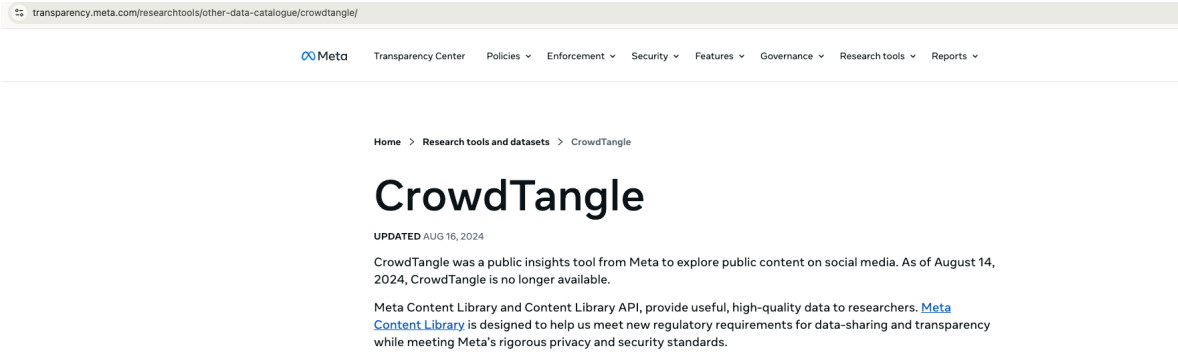
STUDYING SOCIAL MEDIA AFTER THE API CRISIS

17.11.2026

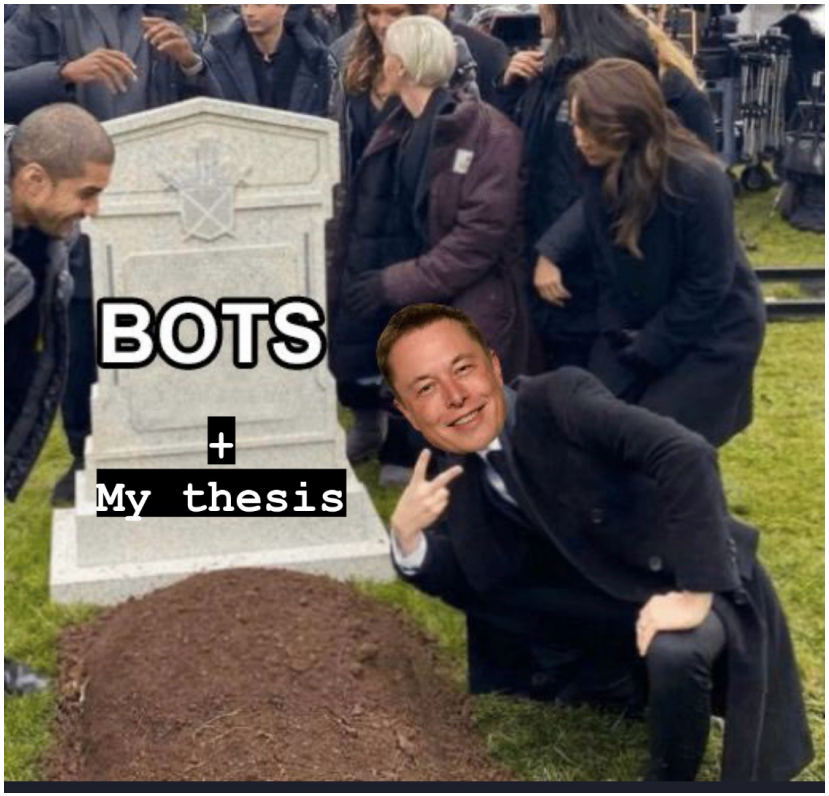
DATA ACCESS IS CLOSING

- Twitter/X ends free API access (2023)
- Reddit restricts its API (2023)
- Meta shuts down CrowdTangle (2024)





DATA ACCESS IS CLOSING



A Celebration of Life and the Untimely Death of CrowdTangle and Other Transparency Tools



Last month, a diverse community gathered to mourn the loss of CrowdTangle, Meta's transparency tool that the company killed on August 14, 2024.

17.11.2026

DATA ACCESS IS CLOSING

Check for updates

BIG DATA & SOCIETY

Original research article

From (almost) open to heavily restricted data access – The development of the Twitter/X developer policies

Big Data & Society
January–March 1–15
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517261419333
journals.sagepub.com/home/bds
Sage

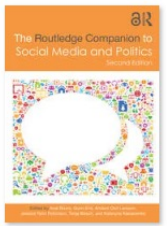
Luisa Golland[1], Oliver Watteler[1], Jonas Recker[1], Jan Schwalbach[1] and Libby Bishop[1]

Abstract

Archiving data is a crucial practice, as it ensures reproducibility of research and aligns with the FAIR principles (Findable, Accessible, Interoperable, and Reusable). To facilitate data sharing, it is essential to understand the legal conditions, especially in the form of policies, at the time of data collection. However, these policies can be difficult to locate, change over time, and may even disappear as new policies are introduced. Terms of Service are a common form of such policies. With the growing use of Twitter/X data in the social sciences, and the restriction of access to the data by the platform itself, investigating ways of sharing of this data has become even more important, and having Terms of Service at hand to check the legal conditions at the time of data collection is a crucial task. To illustrate this use case, we compiled a complete collection of Twitter/X's conditions for sharing data from 2006 until 2023. Using both qualitative and quantitative approaches, we analyse how data sharing conditions have changed. We find that while the Terms of Service become increasingly longer and more complicated, their restrictiveness varies over time. However, our analysis further shows that sharing of full tweet objects both in retrospect and in the future is not usually covered and thus affecting the reproducibility of research as rehydration of tweets is no longer feasible.

Keywords

Open science, social media, Twitter/X, terms and regulations, content analysis, FAIR data



✓ FULL ACCESS

Chapter

So Long Twitter, and Thanks for All the Tweets

By Marco Bastos

Book The Routledge Companion to Social Media and Politics

Edition	2nd Edition
First Published	2026
Imprint	Routledge
Pages	12

Share

ABSTRACT

This chapter reviews the historical contribution of Twitter before it was rebranded as X in July 2023. Twitter was an open platform for social sciences research, particularly political communication, a source of social data so prevalent in the early 21st century that researchers referred to this scholarship as 'Twitter studies'. The chapter revisits the many Application Programming Interfaces that Twitter offered to developers and researchers, including the REST, Search, Streaming, Academic, and Compliance APIs in addition to databases of political communication the company curated and shared with the research community before its contentious acquisition by Elon Musk in late 2022. The chapter concludes with an assessment of the research approaches developed for 'Twitter research' and the extent to which they are transferable to the 'post-API era'.

HOW TO MOVE FORWARD



POLITICAL COMMUNICATION
2023, VOL. 40, NO. 3, 361–366
https://doi.org/10.1080/10584609.2023.2193563

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Farewell to Big Data? Studying Misinformation in Mobile Messaging Applications

Patrícia Rossini

School of Social and Political Sciences, University of Glasgow, Glasgow, UK

KEYWORDS Misinformation; Disinformation; Messaging Applications; Social Media

Concerns about problematic information circulating on digital media platforms have dominated discussions among the public, politicians, and academics alike. Scholars have discussed the conceptual nuances behind different types of false or misleading content, which vary along different levels of falsehood as well as motivation and intent of the sender. For simplicity, I mainly refer to misinformation, i.e., false (or misleading) content shared accidentally, and disinformation, i.e., falsehoods intentionally shared to mislead. Problematic content can also include malinformation – real information used to inflict harm – as well as lies or *bullshit*, with the latter referring to a focus on persuading listeners without any regard for the truth (Carmi et al., 2020; MacKenzie & Bhatt, 2020).

The spread of different kinds of falsehoods during crucial events like the coronavirus pandemic and elections in various countries threatens the normative imperative of a well-informed public as a pillar of democratic citizenship. Messaging applications, however, have been less scrutinized than social media, despite their increasing popularity for news and political engagement. Here, I argue that messaging applications require an important shift in this research agenda, rendering several of the methods used to analyze social media inapplicable. More broadly, these changes also impact research focused on dissonant and divisive digital public spheres.

As platform data close, we should shift our focus from platforms to people (Rossini, 2023)

Move from describing what circulates to explaining how it affects users

Part 2

WHAT IS A CLONE SOCIAL MEDIA ENVIRONMENT?

CLONE SOCIAL MEDIA ENVIRONMENTS

What is a clone?

- A researcher-built website that recreates the key features of a social media feed, used by real participants
- Also called “mock social media”, “social media simulations”, or “experimental research platforms”
- Embedded in an online study: consent → survey (optional) → clone → survey (optional)

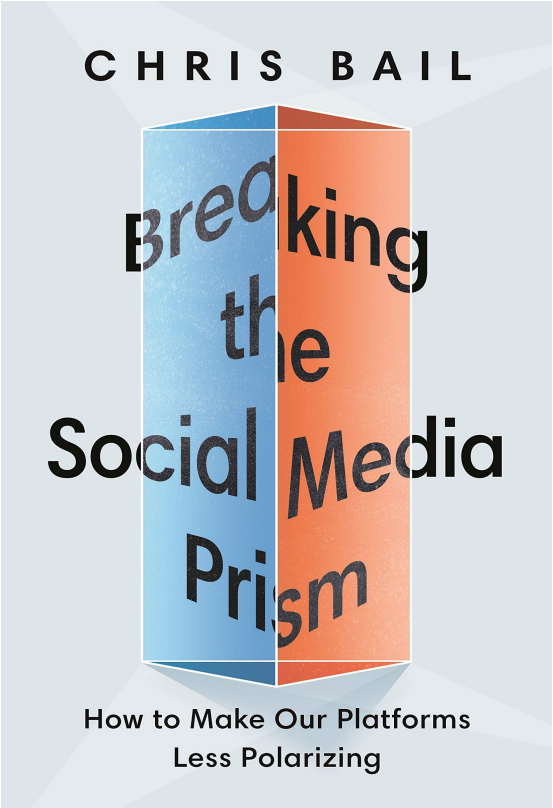
What does it let us control?

- **Content**: which posts exist
- **Exposure**: order, position, repetition
- **Social cues**: like and share counts
- **Measurement**: every click, logged separately

DISCUSSIT

How do different characteristics of social media platforms contribute to political polarisation?

Anonymous conversations about politics with a member of the opposing party, on a politically neutral-looking platform, reduced polarisation (Bail, 2021; Combs et al., 2023)





nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01655-0

Reducing political polarization in the United States with a mobile chat platform

Received: 26 September 2022

Accepted: 14 June 2023

Published online: 21 August 2023

Check for updates

Aidan Combs[1,8], Graham Tierney[2,8], Brian Guay[3,4], Friedolin Merhout[5], Christopher A. Bail[1,6], D. Sunshine Hillygus[6,7] & Alexander Volfovsky[2]

Do anonymous online conversations between people with different political views exacerbate or mitigate partisan polarization? We created a mobile chat platform to study the impact of such discussions. Our study recruited Republicans and Democrats in the United States to complete a survey about their political views. We later randomized them into treatment conditions where they were offered financial incentives to use our platform to discuss a contentious policy issue with an opposing partisan. We found that people who engage in anonymous cross-party conversations about political topics exhibit substantial decreases in polarization compared with a placebo group that wrote an essay using the same conversation prompts. Moreover, these depolarizing effects were correlated with the civility of dialogue between study participants. Our findings demonstrate the potential for well-designed social media platforms to mitigate political polarization and underscore the need for a flexible platform for scientific research on social media.

Political polarization is one of the most pressing social problems of our era[1–5]. Though scholars were once optimistic that the internet could help bridge partisan divides by allowing people to connect with broader communities, many now worry that social media platforms have increased ideological segregation and incivility instead[6–13]. Understanding whether and under what conditions online communication across party lines can shape political divisions is critical to addressing the challenges facing democracy.

Past research provides mixed evidence about the role of cross-party interactions in political polarization. Some studies suggest that these interactions can exacerbate polarization and incivility[14,15]. For instance, recent work concludes that Facebook use may increase political polarization[13], and exposure to ideologically uncongenial information can push partisans further apart[14]. Yet other research suggests that people moderate their views when they engage with those with different perspectives because they come to recognize the value of alternative viewpoints[17–23].

Particularly little is known about the impact of cross-party conversations that occur in online settings[24]. Previous research indicates that online communications are often less civil because people feel less encumbered by the social norms that guide physical interaction[25,26], which can lead them to more easily dehumanize others who disagree with them[27]. This may be especially the case if they are partly or completely anonymous to each other[28–30]. However, there is also evidence that anonymity can encourage people to focus on the content of conversations rather than the identity of the people with whom they engage[31–34]. Anonymity may also allow people to explore alternative viewpoints honestly without fear of social repercussions[15,31,34]. Research is therefore needed to determine whether anonymous cross-party conversations in online settings will exacerbate or mitigate political polarization.

Unfortunately, studying the causal effects of online cross-party conversations presents considerable methodological challenges. Observational analyses of cross-party deliberations on social media

[1]Department of Sociology, Duke University, Durham, NC, USA. [2]Department of Statistical Science, Duke University, Durham, NC, USA. [3]Department of Political Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Department of Sociology, University of Copenhagen, Copenhagen, Denmark. [6]Sanford School of Public Policy, Duke University, Durham, NC, USA. [7]Department of Political Science, Duke University, Durham, NC, USA. [8]These authors contributed equally: Aidan Combs, Graham Tierney. e-mail: alexander.volfovsky@duke.edu

17.11.2026

OUR CLONE: SMCLONE

Lightweight and open source

Built for online studies: links to survey and recruitment platforms through a participant ID

EXAMPLE STUDY: SMCLONE



sociological science

Articles F

Home › Information Diets Are More Diverse in Attention Than in Engagement

Information Diets Are More Diverse in Attention Than in Engagement

Christopher Barrie, Aybuke Atalay, Alia ElKattan

Sociological Science July 21, 2026
10.15195/v13.a33

PDF (1233 views) Reactions 0 Citation

15

ABSTRACT AUTHOR INFORMATION SUPPLEMENTAL MATERIAL PROCESS INFO

What political content do we pay attention to online? Diverse political information is essential for democratic competence, yet online media raises concerns about fragmented information diets. Research on selective exposure highlights how social media can foster ideological echo chambers, while other studies emphasize incidental exposure to diverse viewpoints. A critical measurement challenge is that public platform traces often observe engagement (e.g., likes or shares) more readily than lower-visibility forms of attention—i.e., what users notice or choose to read without necessarily interacting publicly. In this study, we address this gap with a social media clone platform that separately records attention and engagement. We ran the study in both the United Kingdom (once) and the United States (thrice). Across both contexts, we found that the ideology-engagement association is significantly larger than the ideology-attention association. This underscores the importance of measuring attention, rather than solely engagement, to accurately assess the diversity of online information diets.

Do people attend to a wider range of political content than their likes and shares suggest?

4 studies: UK (once) and US (three times)

Recuitment: Prolific and RIWI samples

Design: balanced left–right clone feed · same posts for everyone · random order · 120 seconds

Measures: “Read more” = attention · likes and shares = engagement

Part 3

SETTING UP A CLONE SOCIAL MEDIA ENVIRONMENT

WHAT YOU NEED

1) **Frontend:** shows participants the feed
2) **Backend:** serves the posts to the feed
3) **Data logging and linking:** records every interaction with the feed
4) **Survey tool:** handles consent and questions
5) **Deployment:** puts the environment online for participants

FRONTEND

Runs in any browser, so participants have nothing to install

Can be built with React (for pages and how they behave) and CSS (for how they look)

Design choices to make:

Platform: what platform to “clone”

Content: which posts to show

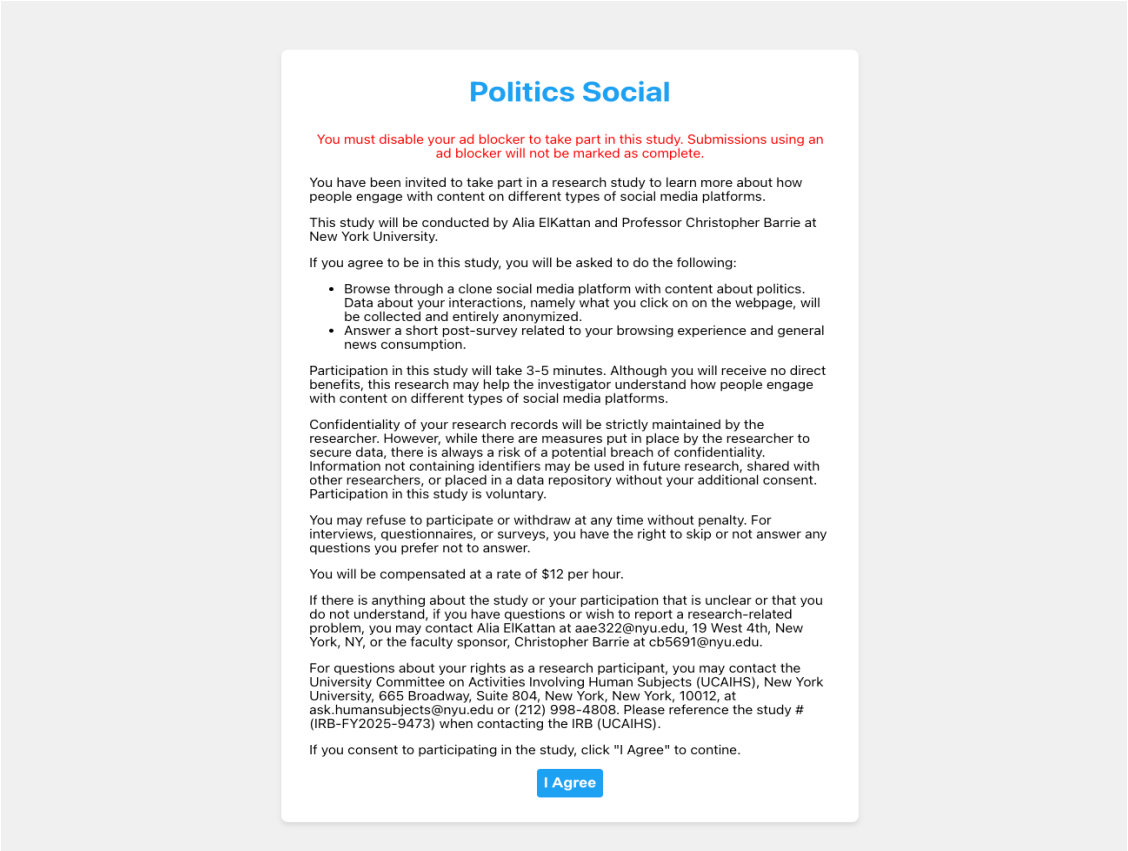
Truncation: how much of each post to show

Duration: how long participants scroll

Engagement: which actions (like, share, comment) to offer and what counts to display

Order: how the posts are ordered

FRONTEND: CONSENT AND STUDY INFORMATION



Politics Social

You must disable your ad blocker to take part in this study. Submissions using an ad blocker will not be marked as complete.

You have been invited to take part in a research study to learn more about how people engage with content on different types of social media platforms.

This study will be conducted by Alia ElKattan and Professor Christopher Barrie at New York University.

If you agree to be in this study, you will be asked to do the following:

- Browse through a clone social media platform with content about politics. Data about your interactions, namely what you click on on the webpage, will be collected and entirely anonymized.
- Answer a short post-survey related to your browsing experience and general news consumption.

Participation in this study will take 3-5 minutes. Although you will receive no direct benefits, this research may help the investigator understand how people engage with content on different types of social media platforms.

Confidentiality of your research records will be strictly maintained by the researcher. However, while there are measures put in place by the researcher to secure data, there is always a risk of a potential breach of confidentiality. Information not containing identifiers may be used in future research, shared with other researchers, or placed in a data repository without your additional consent. Participation in this study is voluntary.

You may refuse to participate or withdraw at any time without penalty. For interviews, questionnaires, or surveys, you have the right to skip or not answer any questions you prefer not to answer.

You will be compensated at a rate of $12 per hour.

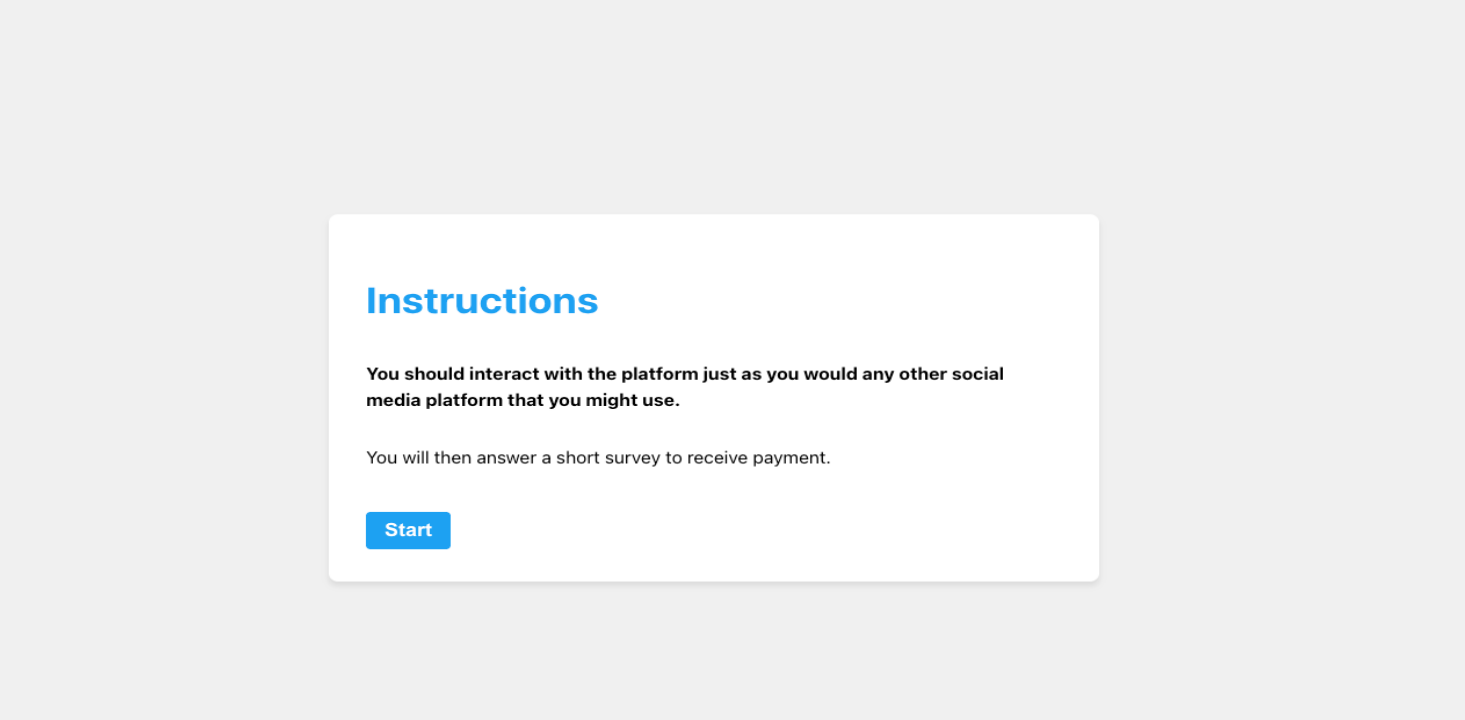
If there is anything about the study or your participation that is unclear or that you do not understand, if you have questions or wish to report a research-related problem, you may contact Alia ElKattan at aae322@nyu.edu, 19 West 4th, New York, NY, or the faculty sponsor, Christopher Barrie at cb5691@nyu.edu.

For questions about your rights as a research participant, you may contact the University Committee on Activities Involving Human Subjects (UCAIHS), New York University, 665 Broadway, Suite 804, New York, New York, 10012, at ask.humansubjects@nyu.edu or (212) 998-4808. Please reference the study # (IRB-FY2025-9473) when contacting the IRB (UCAIHS).

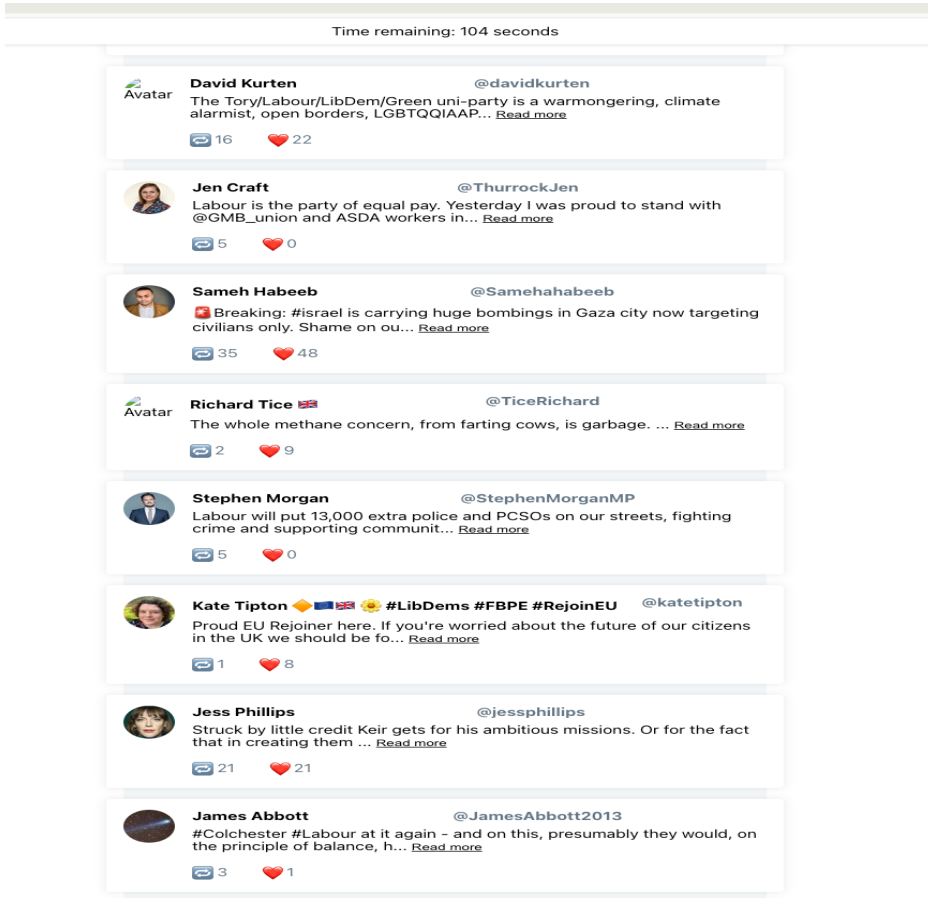
If you consent to participating in the study, click "I Agree" to contine.

I Agree

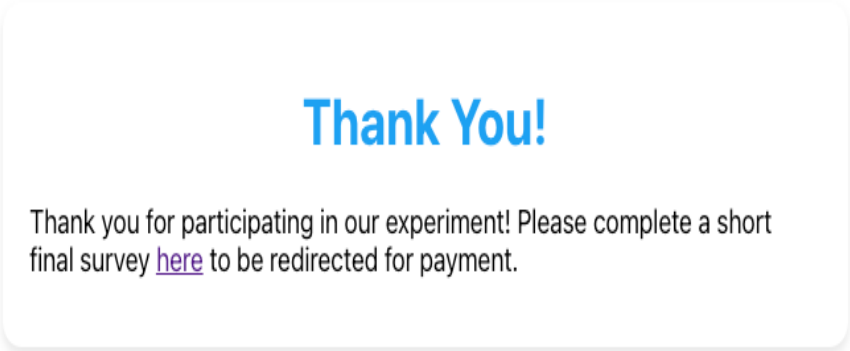
FRONTEND: INSTRUCTIONS



FRONTEND: FEED



FRONTEND: END PAGE



BACKEND

Built with Python (Flask): a small server that runs alongside the frontend

Data choices to make:

Posts: where they are stored (a file or a database)

Conditions: whether everyone gets the same feed or different versions

Order: shuffled, fixed, or set by condition

Images: where they are hosted

DATA LOGGING AND LINKING

Logging: recording what participants do in the feed, with each click saved as an event

In the frontend, each button is linked to a named event:

like → like_click; retweet → retweet_click; “Read more” → expand_tweet

```
52  const handleLikeClick = (tweetId, event) => {
53    event.stopPropagation();
54    toggleLike(tweetId);
55    handleInteraction('like_click', tweetId, userId);
56  };
57
58  const handleRetweetClick = (tweetId, event) => {
59    event.stopPropagation();
60    toggleRetweet(tweetId);
61    handleInteraction('retweet_click', tweetId, userId);
62  };
63
64  const handleExpandClick = (tweetId, event) => {
65    event.stopPropagation();
66    if (!expandedTweets.has(tweetId)) {
67      handleInteraction('expand_tweet', tweetId, userId);
68    }
```

DATA LOGGING AND LINKING

Linking: connecting

- each participant to a unique participant ID
- each event to a participant ID
- each event to a post (through its post ID)
- each participant ID to survey responses

DATA LOGGING AND LINKING VIA POSTHOG

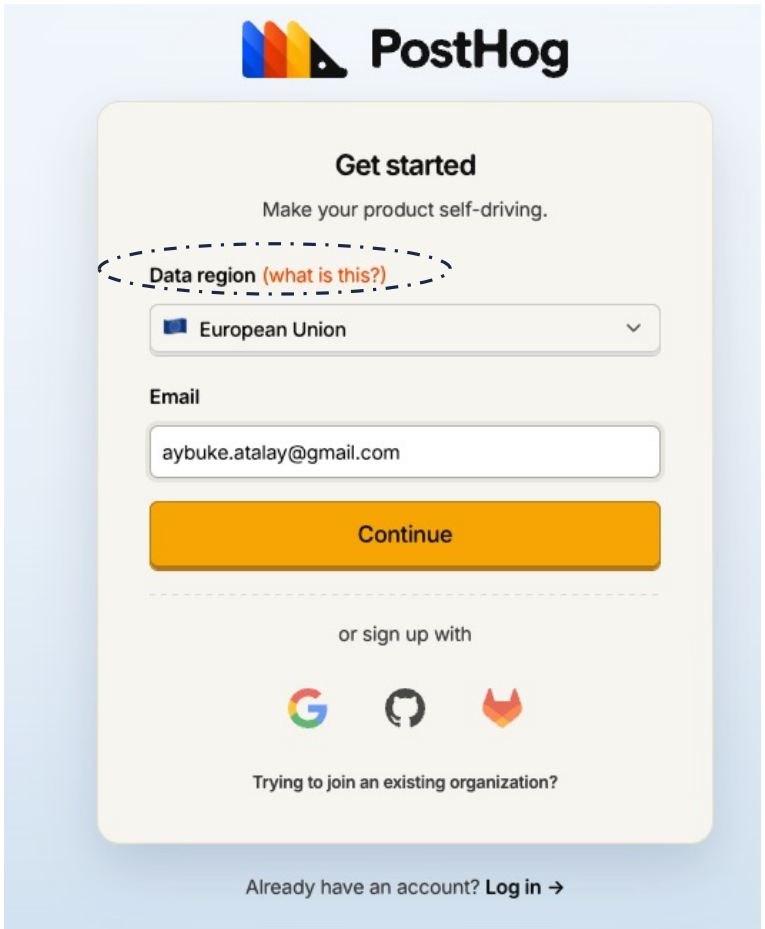
For logging:

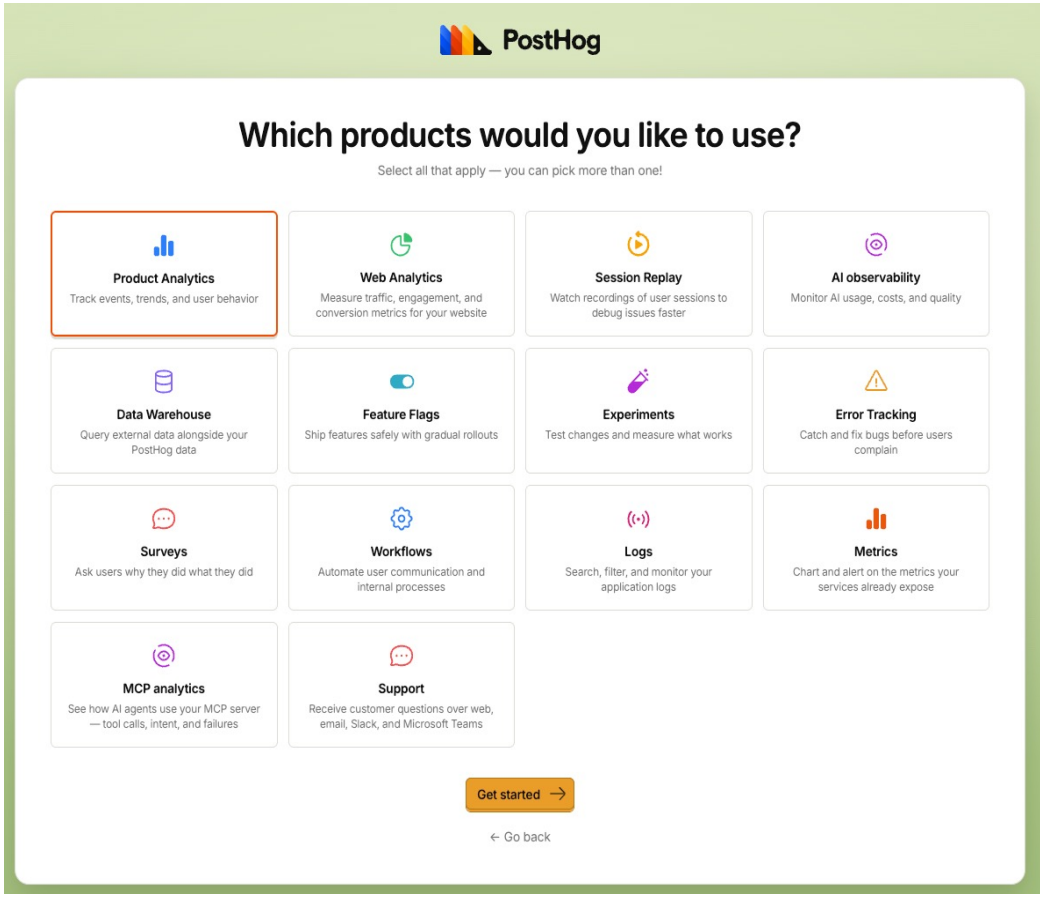
- **Event-level data:** each click recorded with a timestamp
- **Custom events:** you choose what each click records, e.g. the post ID
- **No database needed:** a few lines of code, PostHog stores the rest
- **Live view:** check logging works as clicks arrive

For linking:

- **Participant IDs:** assigned automatically and attached to every click
- **Passed on:** the ID can be added to the survey link
- **Export:** download all clicks, ready to join with survey and post data

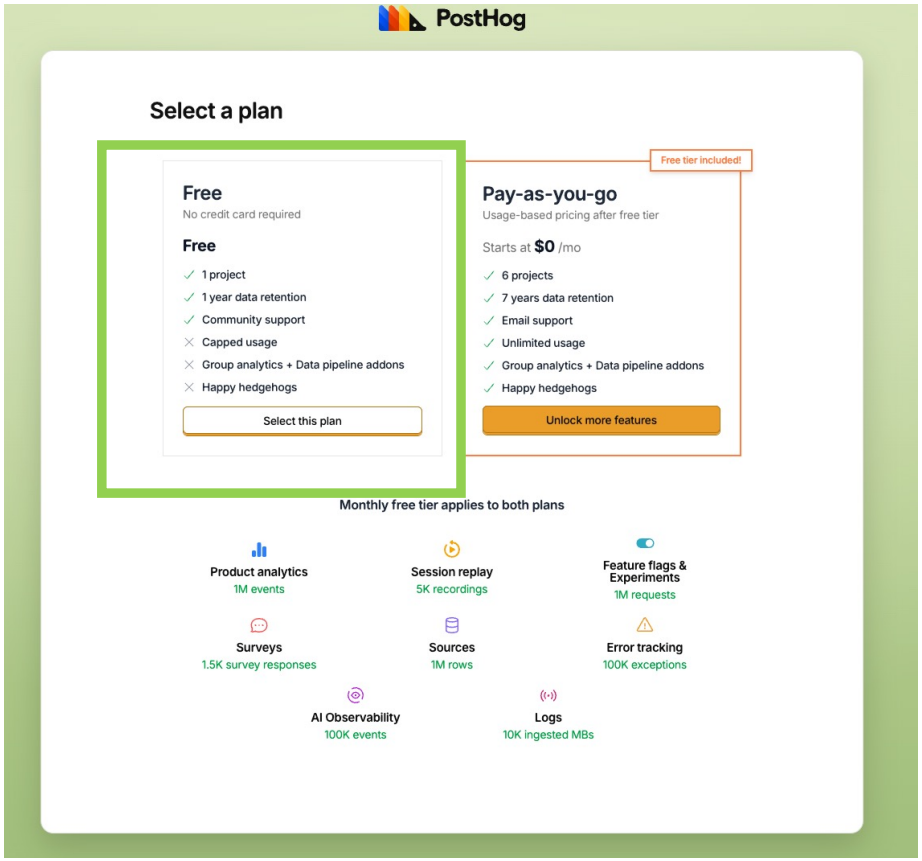
SETTING UP POSTHOG



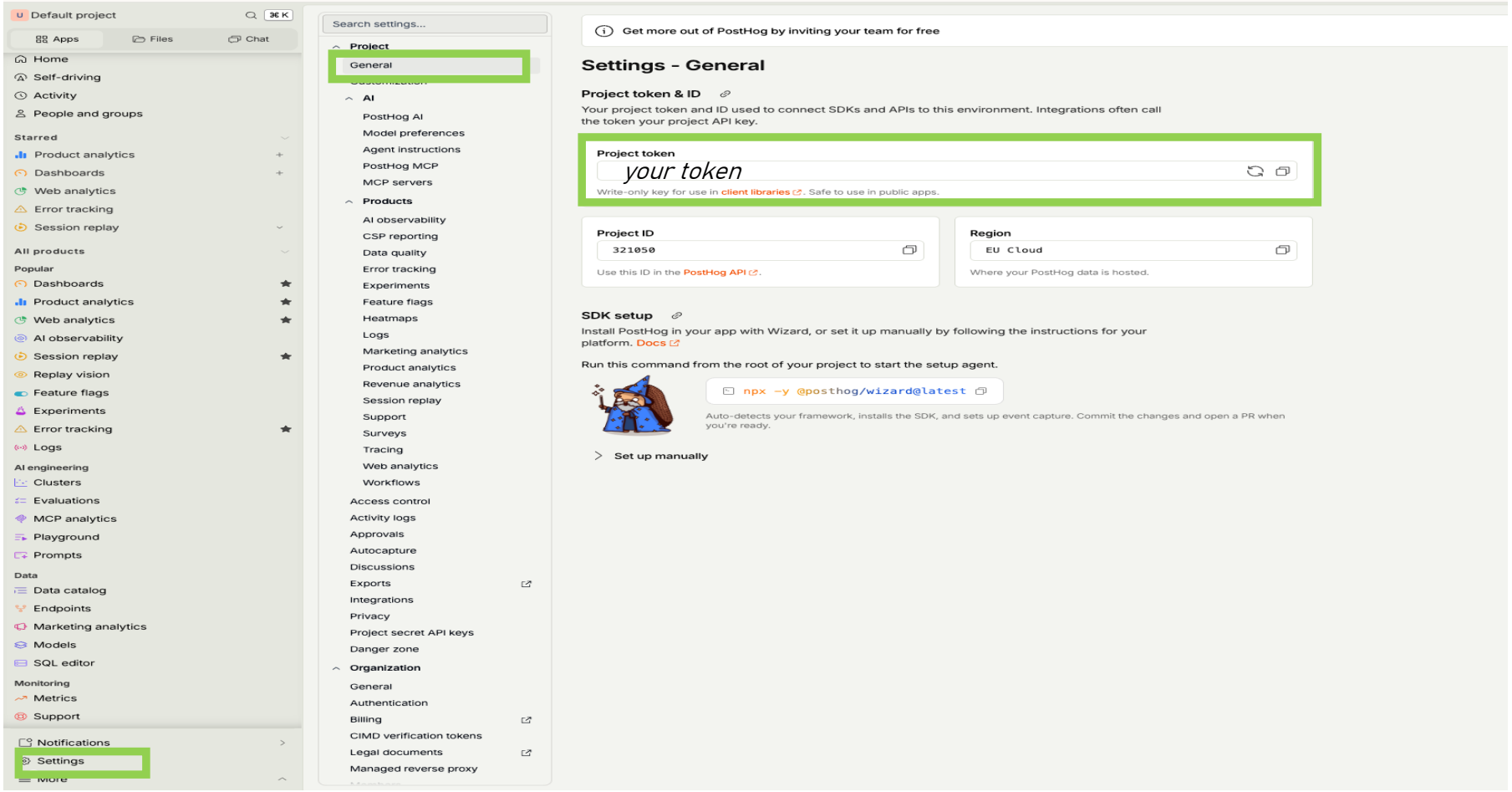


17.11.2026

SETTING UP POSTHOG

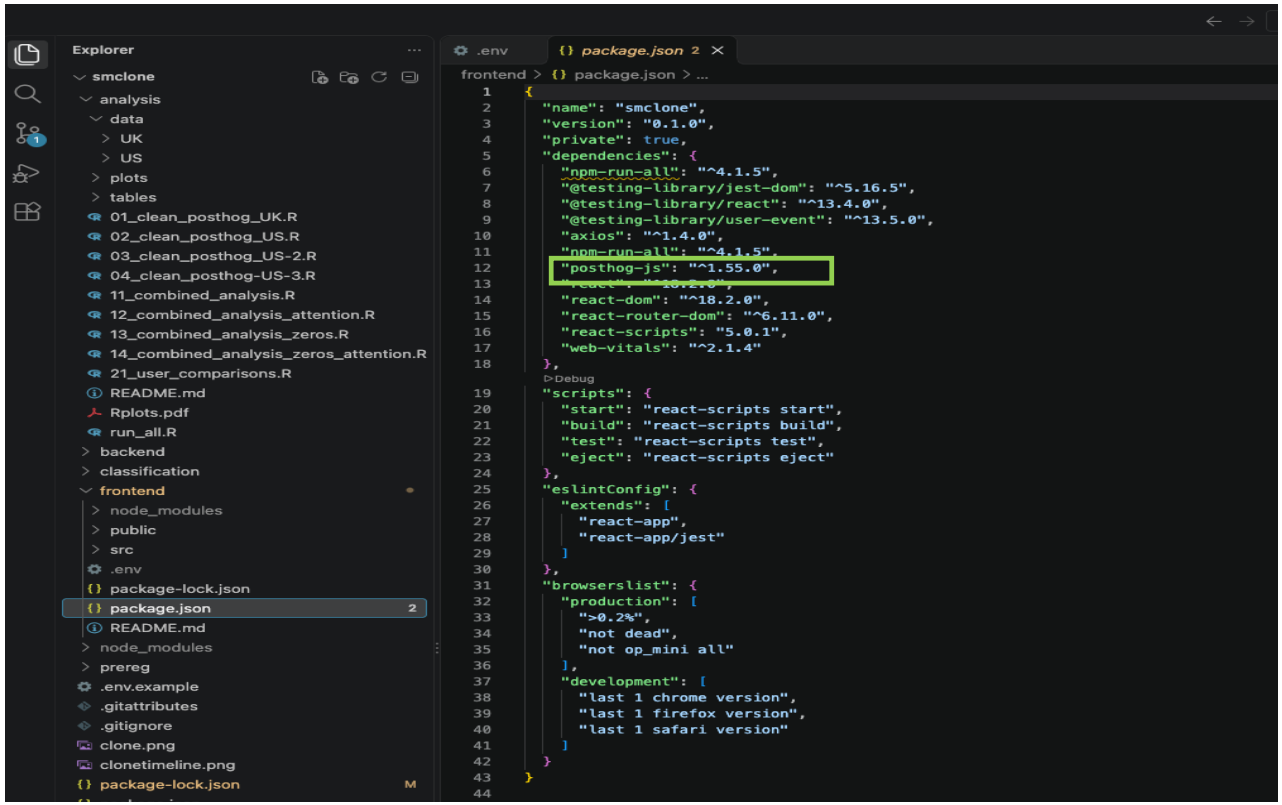


SETTING UP POSTHOG



CONNECTING POSTHOG TO THE CLONE ENVIRONMENT

1. Add PostHog's ready-made code to the clone's frontend



In smclone project:

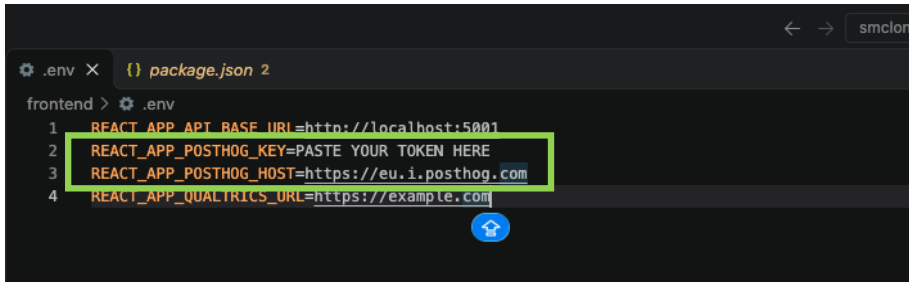
Listed in frontend/package.json,

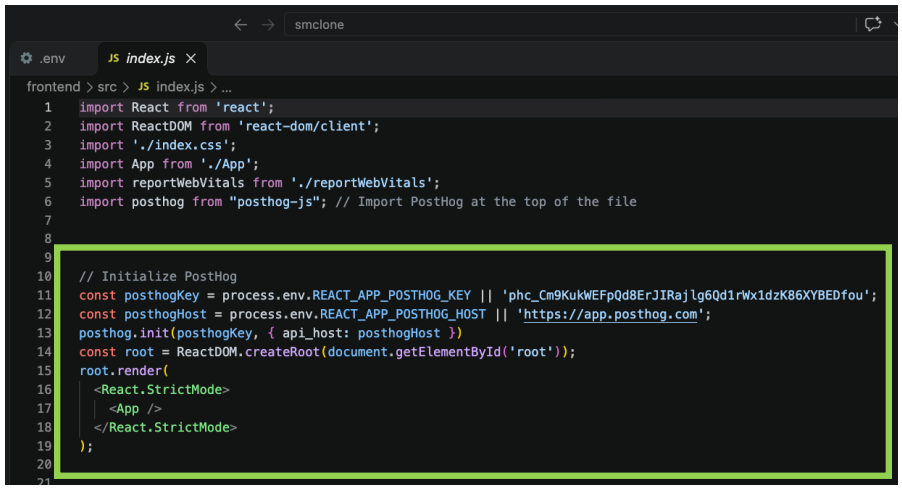
Installed when you run: npm install

17.11.2026

CONNECTING POSTHOG TO THE CLONE ENVIRONMENT

2. Connect the clone to your PostHog project with its token





In smclone project:

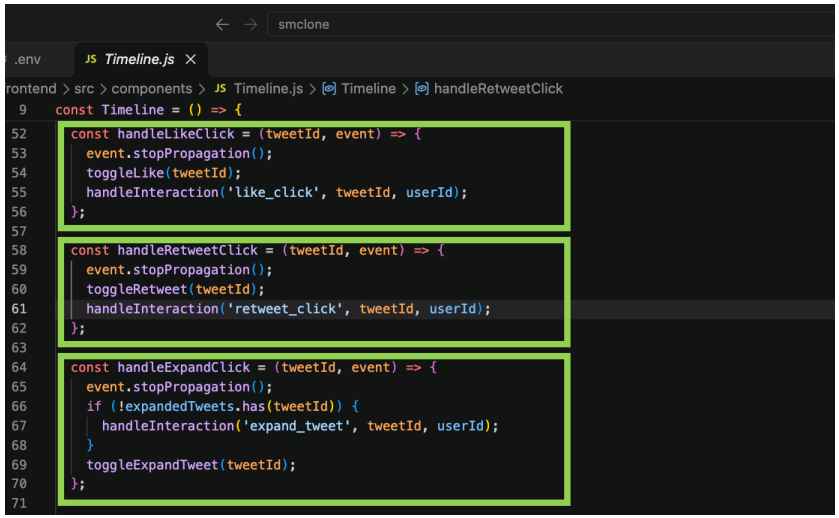
Token saved in frontend/.env (created from .env.example)

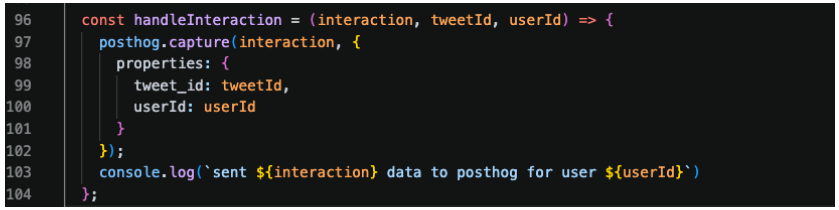
Read by src/index.js when the clone starts

17.11.2026

CONNECTING POSTHOG TO THE CLONE ENVIRONMENT

3. Integrate PostHog into the clone's pages (feed)





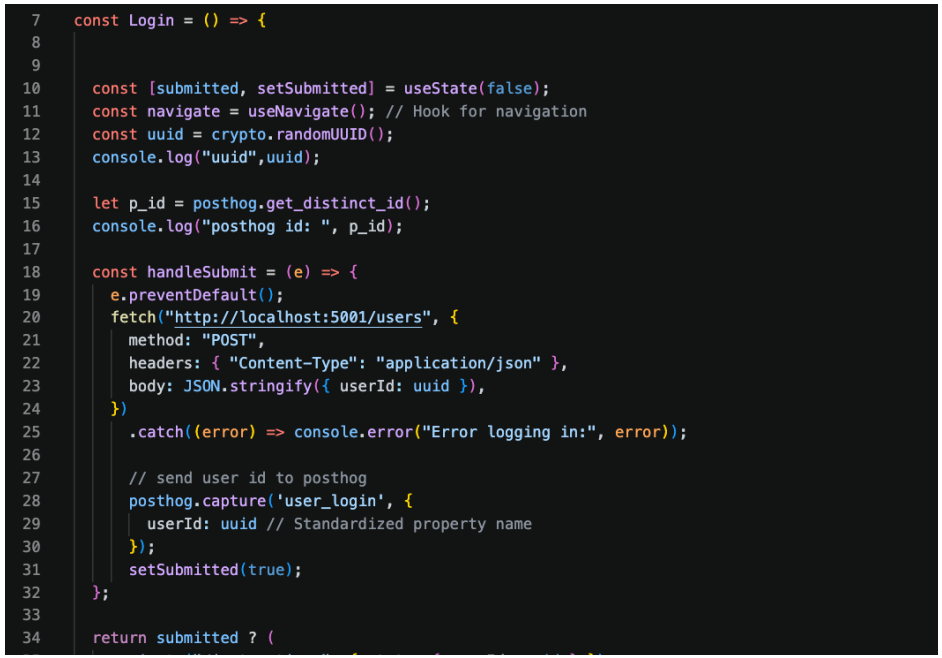
In smclone project:

Each Timeline.js names its event:
like → like_click
retweet → retweet_click
"Read more" → expand_tweet

One function, handleInteraction, sends every event to PostHog with the post ID and participant ID

CONNECTING POSTHOG TO THE CLONE ENVIRONMENT

3. Integrate PostHog into the clone's pages (consent page)



```
const Login = () => {


  const [submitted, setSubmitted] = useState(false);
  const navigate = useNavigate(); // Hook for navigation
  const uuid = crypto.randomUUID();
  console.log("uuid",uuid);

  let p_id = posthog.get_distinct_id();
  console.log("posthog id: ", p_id);

  const handleSubmit = (e) => {
    e.preventDefault();
    fetch("http://localhost:5001/users", {
      method: "POST",
      headers: { "Content-Type": "application/json" },
      body: JSON.stringify({ userId: uuid }),
    })
      .catch((error) => console.error("Error logging in:", error));

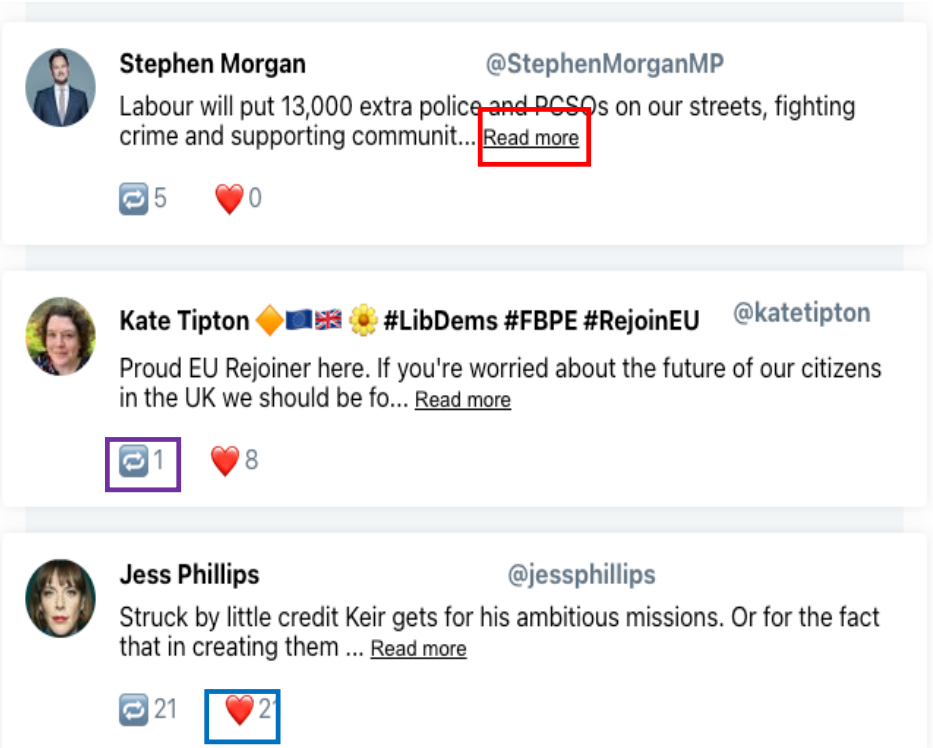
      // send user id to posthog
      posthog.capture('user_login', {
        userId: uuid // Standardized property name
      });
      setSubmitted(true);
  };

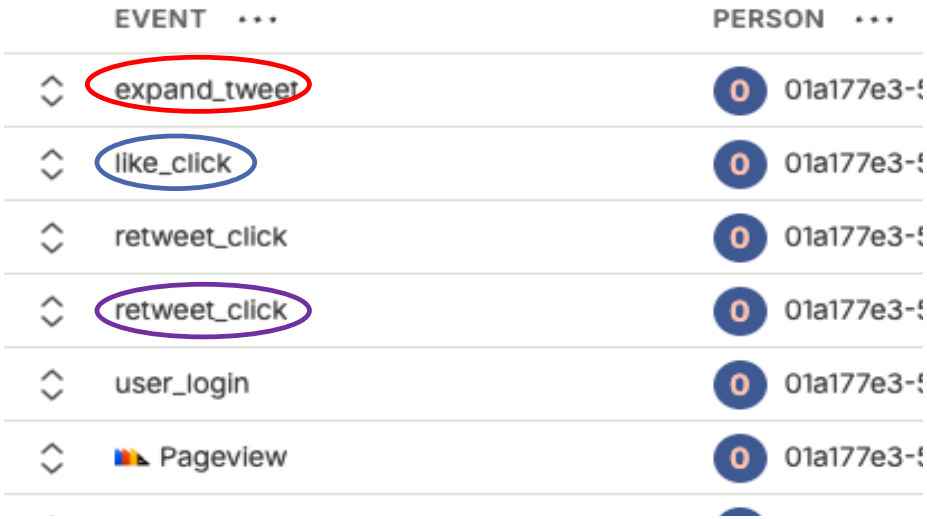
  return submitted ? (
```

In smclone project:

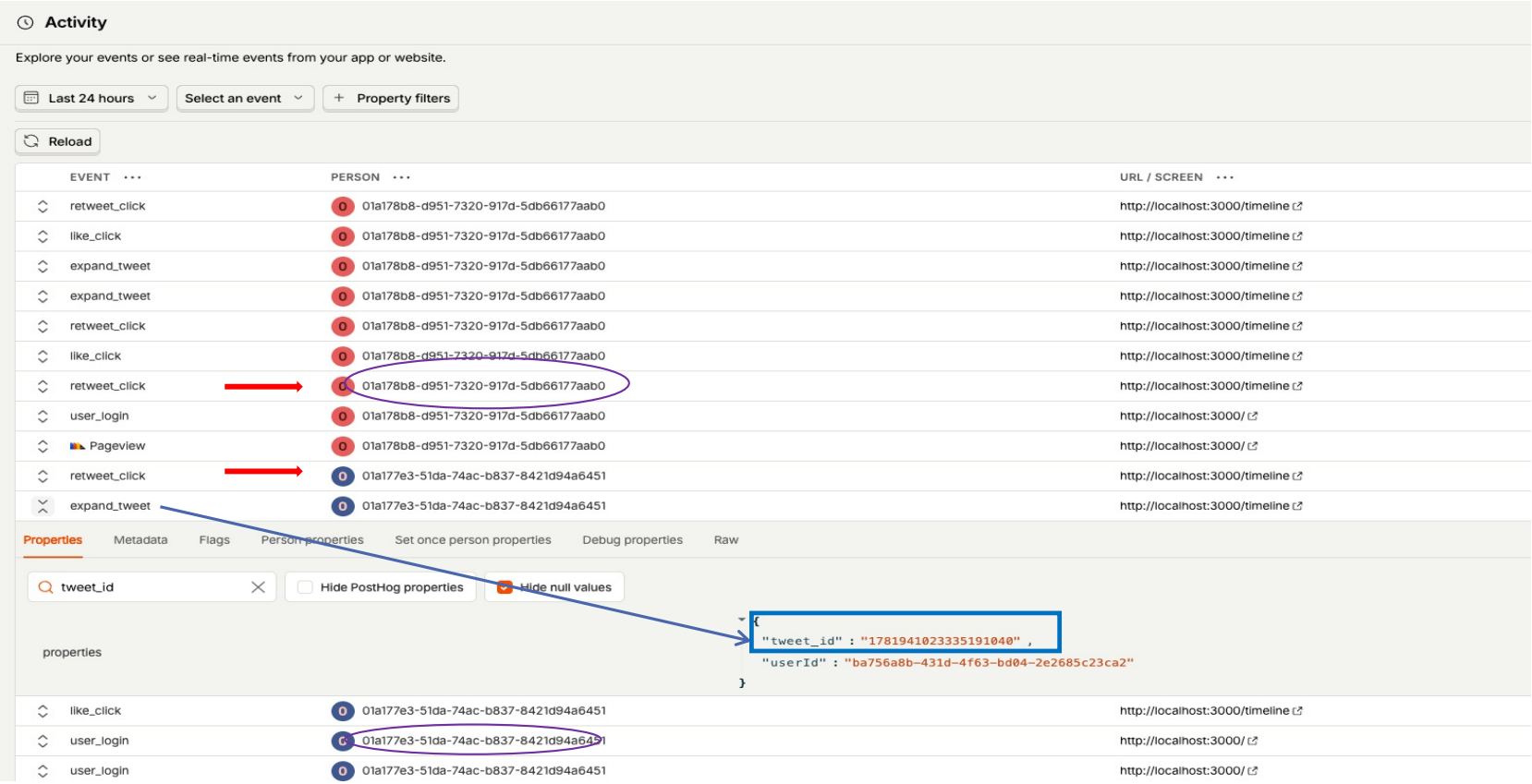
On the consent page (src/components/Login.js), clicking "I Agree" sends a user_login event to PostHog
It marks when each participant started, and carries their participant ID

DATA LOGGING VIA POSTHOG





DATA LINKING VIA POSTHOG



17.11.2026

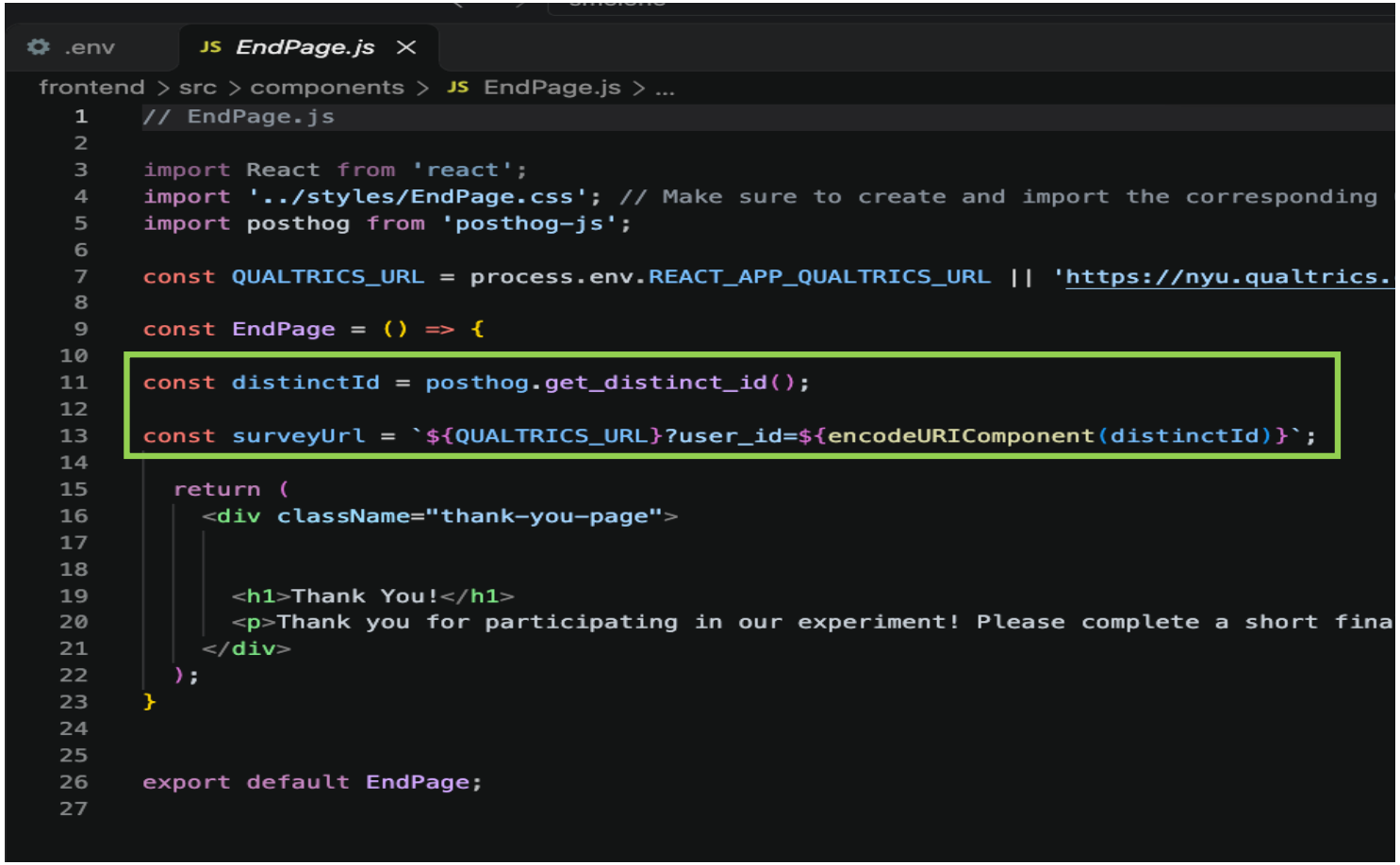
SURVEY

In smclone project:

- The survey comes after the clone: the end page links to it

- Collects measures of ideology, demographics social media habits, and debrief participants

- The link carries the participant ID: …?user_id=<PostHog ID>
- Qualtrics stores the ID in an embedded data field (user_id), so answers can be linked to clicks

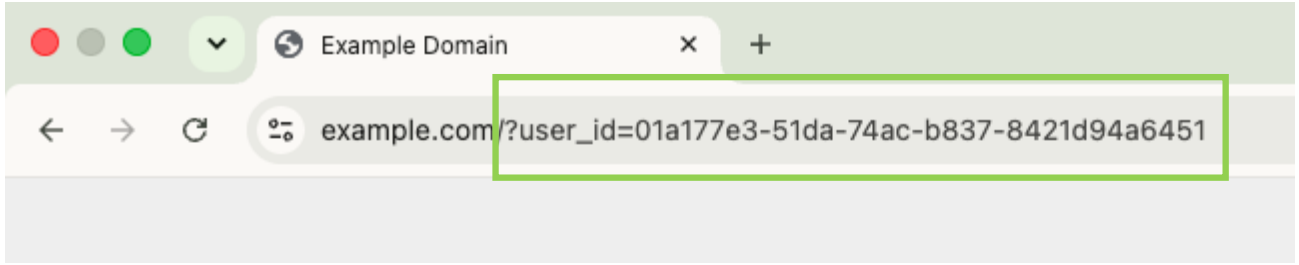

.env
EndPage.js
frontend > src > components > EndPage.js > ...
// EndPage.js
import React from 'react';
import '../styles/EndPage.css'; // Make sure to create and import the corresponding
import posthog from 'posthog-js';
const QUALTRICS_URL = process.env.REACT_APP_QUALTRICS_URL || 'https://nyu.qualtrics.
const EndPage = () => {
const distinctId = posthog.get_distinct_id();
const surveyUrl = `${QUALTRICS_URL}?user_id=${encodeURIComponent(distinctId)}`;
return (

<h1>Thank You!</h1>
<p>Thank you for participating in our experiment! Please complete a short fina

);
}
export default EndPage;

Thank You!

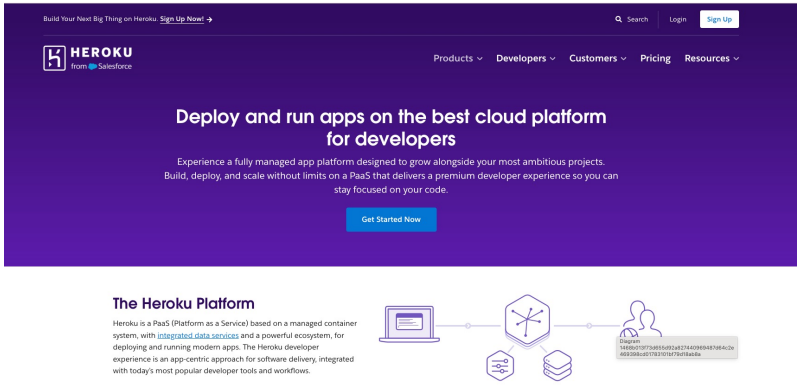
Thank you for participating in our experiment! Please complete a short final survey here to be redirected for payment.



DEPLOYMENT

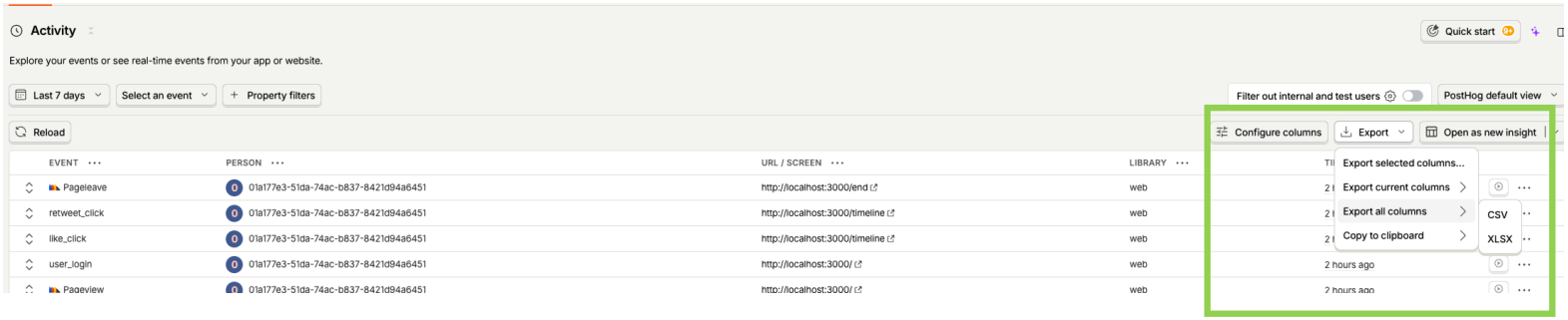
In smclone project: Heroku

- Create a Heroku app and connect it to the GitHub repository
- Enter the settings (PostHog token, survey link, the app's own address) in Heroku's *Config Vars*
- Deployment: Heroku installs the Python and Node packages, builds the frontend (heroku-postbuild in package.json), and starts the server (Procfile)
- The clone is live at https://<app-name>.herokuapp.com, and that's the link participants get, e.g. from Prolific



FROM CLICKS TO DATA

- Each click is an event: participant ID, post ID, action, time
- Export from PostHog and merge with survey responses
- Result: one row per participant per post



GETTING STARTED (LOCAL RUN)

- Code: github.com/cjbarrie/smclone
- Clone the repository
- Install: npm install; pip install flask flask-cors numpy
- Add your PostHog token and survey link to frontend/.env
- Run: flask --app main_uk run --port 5001 and npm start

Live demo

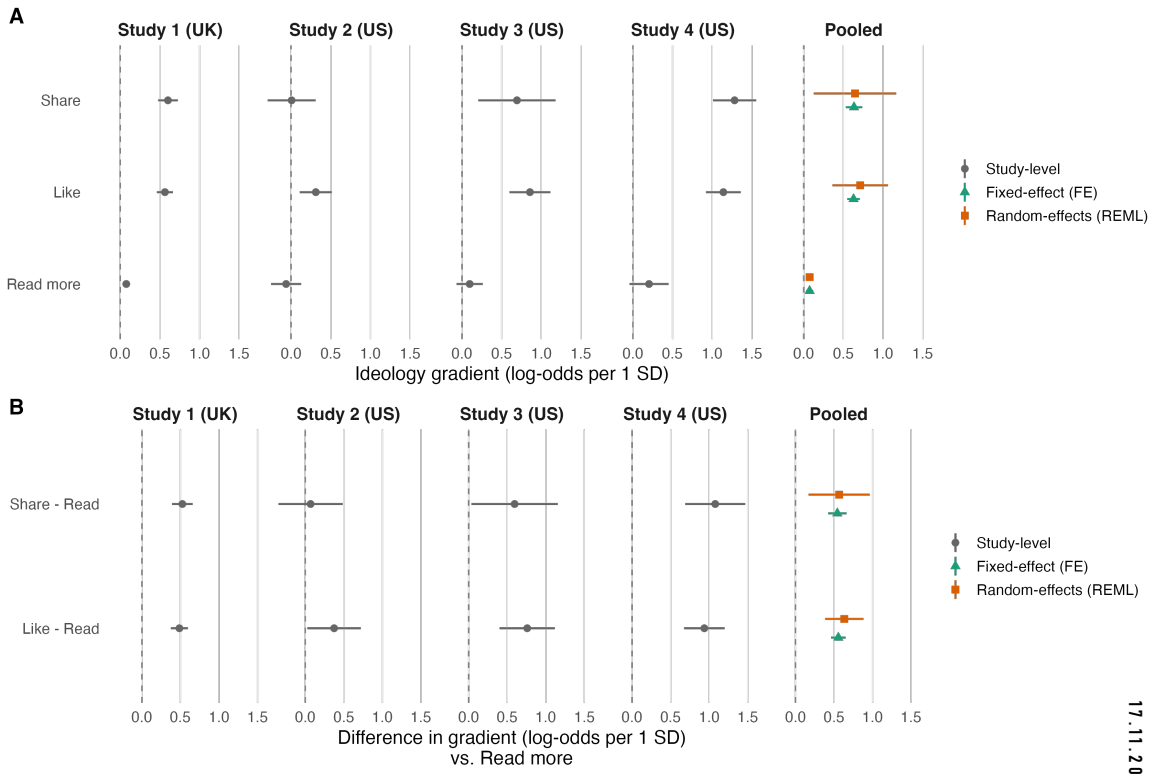
Part 4

WHAT CAN WE MEASURE AND MANIPULATE?

SMCLONE: RESULTS

Attention barely sorts by ideology: left-leaning and right-leaning participants clicked "Read more" on a similar mix of left and right posts. Their attention was fairly balanced.

Engagement sorts strongly: participants mostly picked their own side's posts, indicating their engagement was strongly one-sided.



BEYOND MEASUREMENT: WHAT WE CAN MANIPULATE

- **Content:** which posts appear, e.g. source, topic or ideological balance
- **Order and repetition:** position effects, repeated exposure *(record the order each participant saw)*
- **Engagement counts:** popularity cues *(randomise per participant, not per post)*
- **Labels:** e.g. “AI-generated”, fact-check warnings
- **Framing:** the same claim worded differently

→ *Not limited to these: any feature of the feed can become a treatment*

Part 5

PUSHBACK AND LIMITATIONS

PUSHBACK & LIMITATIONS

Common pushback: “It’s not a real platform”

Limitations

No real stakes: likes and shares reach no one
No real social world: no network, no algorithm
Short exposure: long-term studies are costly

SMCLONE RESOURCES

Code: github.com/cjbarrie/smclone

Paper: Barrie, Atalay & ElKattan (2026), Sociological Science 13: 864–883

Preregistration: osf.io/ny9du

Enquires regarding the paper: Chris Barrie [cb5691@nyu.edu]