CST
PROTECTING OUR JEWISH COMMUNITY

ANTISEMITISM POLICY TRUST

TWITTER:

The extent and nature of antisemitism on Twitter in the UK

CONTENTS

Antisemitism Policy Trust is a registered charity focused on educating and empowering decision makers in the UK and across the world to effectively address antisemitism. The organisation has provided the secretariat to the All-Party Parliamentary Group Against Antisemitism for over a decade.

Community Security Trust (CST) is a UK charity that advises and represents the Jewish community on matters of antisemitism, terrorism, extremism and security. CST received charitable status in 1994 and is recognised by the Government and the Police as a best practice model of a minority-community security organisation.

The Woolf Institute The Woolf Institute is a charitable organisation based in Cambridge. With strong ties to the University of Cambridge, the Woolf Institute is dedicated to research, public education. teaching and policy work with a focus on Christian, Jewish and Muslim communities.

- This report presents an estimate of the level of antisemitism on Twitter in the UK.

- We estimate that there are up to 1,350 explicitly antisemitic tweets in English, newly posted and available to UK users, every day in the UK. A different, second research method, using a different research tool that focuses only on UK-based users, estimated that there are 100 explicitly antisemitic tweets in English from UK-based users every day.

- Based on this estimate, we therefore estimate that there are up to 495,000 explicitly antisemitic tweets per year made viewable for UK users.

- Given the size of the UK's Jewish population (around 290,000 people), there appears to be nearly two antisemitic tweets per year for every Jewish person in the UK.

- Given the difficulty of collecting and analysing Twitter data, it is likely that these figures underestimate the true extent of antisemitism on Twitter in the UK.

- The conclusions are based on the use of two social media data analysis tools (Mozdeh and Pulsar Platform).

- This report marks the end of an 18-month research programme jointly undertaken by the Antisemitism Policy Trust, the Community Security Trust and the Woolf Institute to study the nature and extent of antisemitism online using data from Google, Instagram and Twitter.

INTRODUCTION

Big data is the new lens through which our world is analysed. Crunching hundreds of thousands of lines of code so that we can better understand trends, events and the future is core to many academics, researchers and practitioners across a number of fields. This is no less true in relation to antisemitism, online harm and efforts to make digital spaces safer for users.

As a follow-up to a joint report which analysed the harms occurring through Google's search platform and on far-right web forums, the Antisemitism Policy Trust and Community Security Trust commissioned the Woolf Institute to probe harms occurring on Google, Instagram and Twitter.

This publication represents the third and final installment of that project: the two previous reports showed that Google's Safesearch function is not fit for purpose, and that antisemitism and conspiracism are not only co-located on Instagram but that there is antisemitic supply, rather than demand, on the platform.

Taken together, this series identifies and underlines that anti-Jewish racism persists on major social media platforms, and that either the tools for addressing it are not fit for purpose, or the public policies professed by the world's leading social media giants are at best, poorly enforced and at worst, not worth the paper they are written on.

In the UK, efforts are afoot to regulate social media companies. This legal endeavour is in tandem with complimentary efforts across Europe and beyond. The focus for governments is on the systems that enable harm to be spread, rather than the content. Our reports identify that the current systems – be they Safesearch, algorithms, or automated moderation, are not doing their job. This finding builds on evidence from across the world that, whilst social media can be a force for good, there is little 'safety by design' in an industry designed to disrupt and dismantle.

If antisemitism online and within our societies is to be effectively tackled and reduced, we will need platforms, in this instance acting as agents of harm, to better understand and to act upon what the Big Data is telling them.

BACKGROUND

Using information from a reliable internet source, we estimated that Twitter users worldwide posted, or tweeted, around 290 billion times in 2020. [1]

Based on this figure, we estimated there to be around 790 million tweets per day worldwide.

The last known global estimate from Twitter was made in 2014. They put the number of global tweets per day at 'over 500 million'.[2] This figure suggests our updated estimate for 2020 is reasonably accurate.

Using another reliable internet source, we estimate that the UK represents around 5% of worldwide Twitter usage. In other words, UK users account for around 1 in 20 of all tweets worldwide.[3]

Putting this information together, we estimate that there were around 42 million tweets per day in the UK during 2020. Based on this information, we estimate there to be over 15 billion tweets in the UK that year.

METHODS

We used two resources to collect and analyse Twitter data: Mozdeh and Pulsar Platform. Mozdeh is an open-source social media data analysis tool developed by the University of Wolverhampton.

Pulsar Platform is a commercial social media data analysis tool made available by Fenix Media Ltd (trading as Pulsar).

FINDINGS

PART 1: USING THE MOZDEH SOCIAL MEDIA DATA ANALYSIS TOOL

We began by collecting a large random sample of tweets from Twitter users based in the UK over a 14-day period from 7.21pm 20 July 2020 to 7.21pm 3 August 2020.

Our sample comprised of 22 million tweets. This number represents approximately 1% of all UK tweets in English posted during the period.[4]

These 22 million tweets came from a diverse group of Twitter users and were related to a wide array of discussion topics.

We labelled this group of tweets our 'neutral' dataset. Although our use of Twitter's API meant that Twitter, in effect, selects the tweets to be collected and analyse, we were confident that our sample, given its size, was representative of all tweets sent in the UK around that time.

DISCUSSION OF JEWISH PEOPLE AND THINGS

Our next step was to estimate the number of tweets containing references to Jewish people or things within our 'neutral' dataset. We used Mozdeh to search through the 'neutral' Twitter data using three search terms – 'Jew', 'Jewish' and 'Jews' – and collected tweets with at least one of these terms contained within them. We labelled a group of tweets collected in this as way as our 'Jewish' dataset.

We found 13,144 tweets containing one or more of our three search terms. Although a sizeable number of tweets, this figure represents less than one tenth of one percent of all the tweets that were collected during the research period (0.06%).[5]

It should be noted that the 'Jewish' dataset contained a large number of tweets related to Jewish people and things, but not all.

Tweets that contained related terms (such as, 'kosher', 'synagogue' or 'rabbi') but that did not also contain 'Jew', 'Jewish' or 'Jews' (such as 'rabbi' or 'Chanukah') were not included in our 'Jewish' dataset.[6]

Given Twitter's limit on the number of tweets available when using Mozdeh with the free Twitter API (1%), and if we collected the maximum number of tweets possible in our original 'neutral' sample, then we estimate that there are just under 94,000 tweets per day containing the words 'Jew', 'Jewish' or 'Jews' in the UK.[7]

Readers may be interested to know that, based on these figures, it would take less than three days for there to be more tweets containing the words 'Jew', 'Jewish' or 'Jews' than there are Jewish people in the UK.[8]

4. At the time of this research Twitter was limiting the results from the free Twitter search API to 1% of all tweets. See Thelwall, M. (2015). 'Evaluating the comprehensiveness of Twitter Search API results: A four step method'. Cybermetrics,18-19.

5. This estimate is based on: (13,144 tweets/22,000,000 tweets) x 100.

6. We adopted this strategy to establish a baseline for future comparative work (for example, comparisons with a dataset containing the terms 'Hindi', 'Hindu' and 'Hindus' or 'Muslim' and 'Muslims').

7. This estimate is based on (938.86 tweets per day/1%) x 100 = 93,886 tweets per day.

8. Based on the 2011 Census: 263,346 people who reported their religion as 'Jewish' (NB. The Institute for Jewish Policy Research offer an updated estimate of 290,000. It is available at: https://www.jpr.org.uk/map.)

ANTISEMITISM TERMS FOUND IN THE 'JEWISH' DATA

Next, we wanted to determine the amount of tweets from our 'Jewish' dataset that contained terms associated with antisemitism.[9] This was done to provide clues as to the amount of legitimate discussion of Jewish people and things that might contains antisemitism and to provide the means to make future comparisons of other forms of racial and religious discrimination online.

We took our 'Jewish' dataset (tweets containing the search terms 'Jew', 'Jewish' and 'Jews') and used the Mozdeh social media data analysis tool to search through it using a list of keywords and phrases commonly associated with antisemitism.

Our list of search terms included antisemitic phrases such as 'Jewish filth', 'Jewish lobby', and 'Jewish Nazis'. Other terms included those that are not explicitly antisemitic but are frequently used alongside more antisemitic phrases.

We labelled all these terms as our 'antisemitism' terms (i.e. terms frequently associated with antisemitism). They are all listed in an appendix to this report.

We found 1,050 tweets in our 'Jewish dataset' that contained one or more of our 'antisemitism' terms. This number represents around 8% of all tweets from our 'Jewish' dataset. In other words, around 1 in 12 of all tweets containing the terms 'Jew', 'Jewish' or 'Jews' also contains a term frequently associated with antisemitism.

1,050 tweets over 14 days is the equivalent of 75 tweets per day. Remember, we estimated that we had around 1% of all tweets in English visible in the UK. Therefore, we estimate there to be 7,500 tweets per day containing 'Jew' 'Jewish' and 'Jews' and one of our antisemitism terms.[10]

9. Again, we adopted this strategy to facilitate future collaborative work (e.g. a comparison with a dataset containing the terms 'Muslim and 'Muslims').

10. This estimate is based on: 1,050 tweets/14 days = 75 tweets per day; (75 per tweets per day)/1% x 100% = 7,500.

BACK TO THE 'NEUTRAL' DATASET

Having completed the task of estimating the proportion of tweets from the 'Jewish' dataset that contain an antisemitism term (for the purposes of future comparative work), we turned our focus to an estimation of antisemitism across the whole of Twitter in the UK.

Assuming that there would be many tweets that not did not contain the terms 'Jew', 'Jewish' or 'Jews' but that did contain one of our antisemitism terms, we returned to our original sample of general Twitter data (our 'neutral' dataset).

This time, rather than estimate the proportion of Twitter conversations related to Jewish people containing terms associated with antisemitism, our aims were to estimate the amount of antisemitism across the whole of Twitter in the UK.

From our original sweep of Twitter (the 'neutral' dataset), we found nearly 2,000 tweets containing one or more of our antisemitism terms (1,899). Whilst this is less than one hundredth of one percent of all tweets posted in the UK during our time period, the figure is equivalent to 13,564 tweets per day, or the equivalent of nearly 5 million viewable tweets per year in the UK.[11]

11. This estimate is based on: 1,899 tweets /14 days = 135.64 (rounded up/down); (135.64/1 x 100) = 13,564 tweets per day or 4,954,355 tweets per year. (NB. Per year figure based on 13,564 tweets per day x 365.25 days.)

IDENTIFYING EXPLICIT ANTISEMITISM ON TWITTER

12. Krippendorff alpha reliability score of 0.64, where 1 is perfect agreement. Scores over .80 are normally used as a benchmark of reliability. As stated, our score reflects the difficulty of the task.

Since not all of our antisemitism search terms were explicitly antisemitic, and given the tweets containing antisemitism terms may be used in non-antisemitic ways (e.g. identifying, calling out or discussing antisemitism, or quoting antisemitic text to highlight discrimination), we examined the tweets for explicit expressions of antisemitism.

From the 'neutral dataset', we selected at random 320 tweets containing one of our antisemitism terms. To determine the level of antisemitic content in the tweets, we asked experts from APT and CST plus a third non-expert to annotate (i.e. review and score) the images using a three-point 'traffic light' system: 'yes' for obvious antisemitism; 'maybe' for borderline cases (or where the annotator was undecided); and 'no' for no obvious antisemitism.

We used a reliability statistic to measure the level of agreement between our three annotators. Overall, our reliability score reflected the difficulty of the task although we found at least some level of agreement between the annotators. [12]

To produce an overall score for a tweet, we used the majority vote of the three annotators (which is also the mode of the scores). The annotator accuracy with respect to the majority vote was 75%, 98%, and 96%.

The classification of tweets as antisemitic or not, as measured by a majority vote among the annotators, was as follows:

- 10% of the tweets were deemed to be antisemitic (33 tweets out of 320);

- 31% of the tweets were classified as borderline or undecided (98 tweets out of 320);

- 59% of the tweets were deemed not to be antisemitic (189 tweets out of 320).

13. Our exact estimate of 1,356 antisemitic tweets per day equals 495,435 such tweets per year (reported in the introduction as 495,000).
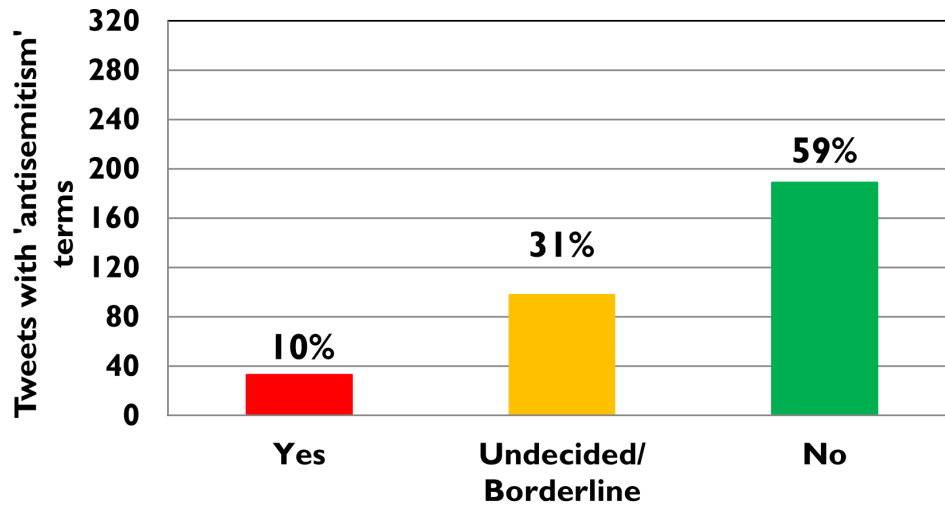
Therefore, if we assume that 10% of tweets containing terms frequently associated with antisemitism are likely to be explicitly antisemitic, we can estimate – using our figure of 13,564 tweets per day containing such terms – that there are 1,356 tweets per day or just over 495,000 tweets per year containing explicit antisemitism.[13]

These figures represent roughly one antisemitic tweet for every two Jewish people in the UK.

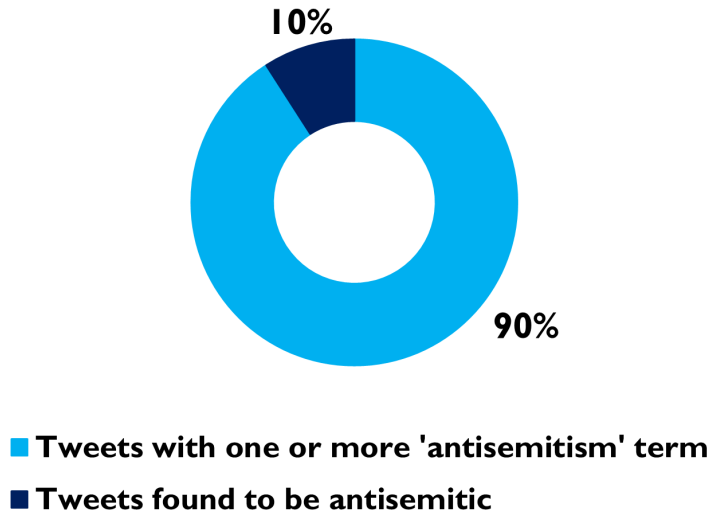**Fig.1 Results of annotation of tweets containing 'antisemitism' search terms**

Yes = Antisemitic, Undecided/Borderline = Don't know, No = Not antisemitic

Fig.2 Proportion of tweets with one or more 'antisemitism' term found to be antisemitic

■ **Tweets with one or more 'antisemitism' term**
■ **Tweets found to be antisemitic**

FINDINGS

PART 2: USING THE PULSAR PLATFORM SOCIAL MEDIA DATA ANALYSIS TOOL

14. Our estimate is based on: 60,791 tweets/56.3 days = 1,080 tweets per day. We used 56.3 days to account for the exact timings between 8.07am on 17 July 2020 and 2.30pm on 11 September 2020

15. Rounded down to 100 in the introduction.

We used the Pulsar Platform tool to further explore the nature and scale of antisemitism on Twitter.

We worked closely with experts at Pulsar and conducted a second, much larger search of Twitter. We used the same antisemitism terms used for the search using the Mozdeh tool and collected tweets containing one or more of these terms for an 8-week period between 17 July and 11 September 2020.

Using our search terms, we collected 60,791 tweets. Assuming that Pulsar Platform collected all tweets matching our search terms, our dataset of tweets represented a tiny fraction of the UK tweets sent during the search period (0.003%).

Despite the low percentage, the figure represents just over 1,000 tweets per day (1,080 tweets).[14]

Based on the annotation of the data collected using Mozdeh, we might assume that 10% of the tweets collected by Pulsar Platform contain explicit expressions of antisemitism. This would give us a figure of 108 explicitly antisemitic tweets per day or 39,420 antisemitic tweets per year.[15]

EXPLAINING THE DIFFERENCES BETWEEN MOZDEH AND PULSAR

The use of Mozdeh and Pulsar Platform returned different levels of explicit antisemitism: 100 explicitly antisemitic tweets in English per day with Pulsar; 1,356 explicitly antisemitic tweets in English per day with Mozdeh. What might explain this discrepancy?

Access to social media is often mediated by commercial companies. This presents something of a challenge for researchers. The benefits of access must be weighed up against the lack of disclosure around precise data collection methods which remain, in effect, trade secrets (in this case trade secrets held by Twitter and Pulsar).

Pulsar Platform uses proprietary methods which are not revealed in full to clients or third parties. This means the authors do not have a full knowledge concerning the criteria with which data were collected.

Two facts are known. Pulsar results are tweets by UK-based Twitter users. On the other hand, Mozdeh results are tweets in English made available to UK-based Twitter users, including those from elsewhere in the world.

Also, it should be noted that Mozdeh offer a '1% stream' (i.e. access to 1% of all tweets related to search terms), whereas Pulsar offer a '100% stream' (i.e. access, it is claimed, to all tweets related to search terms).[16]

An implication of all of this is the variance between findings, in our case the discrepancy between the daily amount of tweets containing antisemitism identified by Mozdeh and Pulsar. In effect, the discrepancy offers a range of prevalence. Taking conservative and common sense approaches, using both methods established a range of search results from Pulsar's lower estimate to Mozdeh's higher estimate.

The findings reported in this report offer estimates based on the most responsible use of Mozdeh and Pulsar. They are offered in the spirit of scientific enquiry and the research team would welcome a refinement of these methods and the generation of more accurate findings by colleagues in the field.

The aim here is to offer an estimate of the number of antisemitic tweets per day and per year on Twitter in the UK. Our estimates are, in effect, initial baseline frequency or prevalence measures (of tweets per day and year) against which future measurements may be compared.

16. Our estimate is based on an assumption that we collected all of the 1% Mozdeh stream and assumes Twitter supply all relevant tweets to that stream.

We found that the number of tweets collected using Mozdeh is equivalent to around 20% of global tweets, which is in line with estimates of the proportion of global tweets that are in English.

Our 20% figure is based on: (22 million tweets in English)/14 * 365 = 576 million tweets per year. Scaling that up from 1% to 100% gives 57.6 billion tweets.

Recall that our estimate of global tweets was 290 billion for the year 2020. Based on all this, 57.6 billion tweets in English represents 19.8% of global tweets. In 2018, authors writing on web analysis platform Vicinitas estimated that 32% of tweets were in English (see https://www.vicinitas.io/blog/twitter-social-media-strategy-2018-research-100-million-tweets#language).

DESCRIPTION OF TWITTER USERS AND TWEETS

Using Pulsar Platform we were able to identify biographical and geographical characteristics of Twitter users posting tweets which matched our antisemitism terms.

Please note, the information below is intended to offer biographical and geographical details of Twitter users engaged in conversations featuring terms frequently associated with antisemitism.

We make no claims whatsoever about whether or not these conversations or accounts feature explicit expressions of antisemitism. This is not intended as a measure of Twitter accounts that are actively antisemitic.

Gender.

- Male = 34%
- Female = 19%
- Unknown = 48%

(% rounded up)

Locations Top Ten:

- London
- Manchester
- Redcar
- Liverpool
- Glasgow
- Brighton
- North West Leicestershire
- Oxford
- Birmingham
- City of Edinburgh

Bio Keywords Top 10

- socialist
- politics
- music
- labour
- proud
- member
- writer
- people
- party
- supporter

Top 10 topics

- Israel (micro-blog count [individual Twitter accounts) = 12,786)
- racism (micro-blog count = 6,558)
- Jews (micro-blog count = 6,430)
- Palestinians (micro-blog count = 4,037)
- people   `
- Gaza (micro-blog count = 2,849)
- apartheid Israel (micro-blog count = 2,710)
- power (micro-blog count = 2,411)
- apartheid (micro-blog count = 2,386)
- antisemitism (micro-blog count = 2,376)

As stated, we make no claims whatsoever about whether or not these accounts contain explicit expressions of antisemitism. Rather, these are the discussion topics, as identified and labelled by Twitter, in which terms associated with antisemitism are most likely to be found on the platform.

APPENDIX

Antisemitism search terms

Search query for posts containing 'antisemitism terms' (i.e. terms associated frequently with antisemitism) in English and from the UK:

(LANG en) AND (LOCATION GB) AND

((jew OR Jewish OR Israel OR jews OR judaism OR zionism OR zionist) AND ((nazi OR nazis) AND (Palestine OR gaza OR Palestinian OR Palestinians OR IDF OR "defence force") OR Rothschild OR Soros OR Epstein OR filth OR coronavirus OR BLM OR 'black lives matter' OR disloyal OR disloyalty OR "global affairs" OR power OR control OR banking OR banks OR finance OR globalisation OR hollywood OR executives OR globalists OR globalist OR globalise OR racism OR '9 11' OR discrimination OR "death of jesus" OR "killed jesus" OR (State OR government OR politics OR political) AND (control OR power OR manipulation OR hypnotising) OR Greedy OR "blood libel" OR "child blood"~3 OR "children blood"~3) OR holocaust AND (hoax OR myth OR fake) OR "jewish lobby" OR Israel AND Apartheid OR synagogue AND satan)