

COVID Tracker: Artificial Intelligence (AI)-powered dashboard for COVID-19 related public sentiment and opinion mining in social media platforms

AIMS

We aimed to create an artificial intelligence (AI)-driven data dashboard for policymakers, and public health and clinical practitioners to aid in understanding and mitigating the impacts of the COVID-19 pandemic. To achieve this, we sought to:

- Develop the world's first dashboard prototype to enable in-depth analysis and visualisation of public sentiment on two popular and most representative social media platforms, namely Facebook and Twitter, in relation to key COVID-19 issues of interest to policymakers
- Power the dashboard by a first-of-its-kind 'human expert-in-the-loop' (HITL) AI tool, that incorporates health and public policy expert knowledge, to identify relevant comments, categorise these into key COVID-19 themes, and analyse the sentiments expressed.
- Develop a custom validation tool for continuous training and optimisation of the AI algorithms in order to ensure the highest quality and reliability of results.

KEY FINDINGS

- A framework of key COVID-19 themes was designed following interviews with public health policy experts and Scottish Government (SG) representatives.
- A new benchmark dataset of millions of Twitter and Facebook postings created by UK users was collected and shared for further study.
- A novel prototype dashboard was developed and demonstrated to key stakeholders. This integrated, for the first time, simultaneous Facebook and Twitter data analyses for user-selected COVID-19 themes.
- A set of innovative tools were created to help continuously optimise and refine HITL AI models for theme and sentiment classification. These incorporated real-time predictions and a new evaluation process to automate learning from subject matter experts.
- Exemplar case studies were developed to evaluate and validate our AI models and dashboard prototype, specifically: automated analysis of a recent SG Public Dialogue dataset; understanding UK public attitudes on social media towards COVID-19 vaccination and contact tracing apps; assessing impact of COVID-19 on ethnic minorities; and pilot analysis of long-COVID discourse on social media and online support groups.
- Our case studies results showed that AI-based methods can complement findings using conventional methods (such as surveys) to monitor public sentiment whilst providing greater spatio-temporal granularity in near real-time. We discovered that sentiment expressed via Facebook tended to be more positive, in general, than via Twitter, including when studying the discourse around effects of COVID-19 on minority ethnic groups.

WHAT DID THE STUDY INVOLVE?

- A requirements analysis process was carried out at the start of the project in order to identify user needs and determine system requirements for the dashboard. This involved interviews and focus groups with key stakeholders, including public health policy experts and SG representatives. The feedback and expert advice from stakeholders was synthesised into 24 key themes of interest, with over 350 sets of keywords identified as marking pandemic-related ideas.
- A thorough privacy assessment was conducted involving consultation with multiple governance bodies before determining the minimum level of information necessary to complete our public interest goals. It was assessed that using public posts, publicly shared location information and metadata, such as time, date and public reactions, was of minimal risk to privacy.
- In order to capture the most relevant conversations about COVID-19, we chose Facebook and Twitter platforms. We considered alternate networks such as Reddit and Instagram, but limited our scope to the top two, since they provide quick feedback to ongoing changes in health policy, as well as being the most popular and most representative social networks currently in use. We recognise this dataset may not represent a completely unbiased reading of events, however as evidenced by findings from our exemplar case studies, it provides a useful expression of general public attitudes and response to crises, and can also serve to point out areas of concern, such as the spread of inaccurate or misleading information.
- We were able to use Twitter's streaming interface to collect local posts following its code of ethics and privacy obligations. In this sense, our infrastructure receives new posts marked with locations within the UK continuously and nearly instantaneously. We also partnered with Facebook's CrowdTangle service, which provides access to publicly available pages, groups and posts across the platform. The tool allowed us to identify pages and groups relevant to our key themes, from which data was collected for further analysis.
- The collected social media data were analysed using state-of-the-art Natural Language Processing (NLP) tools to determine their emotive content—whether the author was communicating positive or negative sentiment. The 'overall' polarity of each social media (Twitter/Facebook) post was predicted and randomly selected predictions were validated by human experts, before being used to develop, train and test machine learning-based NLP algorithms and models for topic and theme classification. The models were evaluated by investigating public sentiments around the 24 key COVID-19 themes, which included, amongst other topics, Education, Social/Family, Business Restrictions, and Mental Health and Loneliness.
- Further evaluation was conducted through priority case studies to analyse: the SG Public Dialogue consultation on COVID-19 response; the impact of the pandemic on ethnic minorities; and public attitudes towards COVID-19 vaccinations and contact tracing apps. A pilot study was also carried out on 'long-COVID' topics.

WHAT WERE THE RESULTS AND WHAT DO THEY MEAN?

- A prototype COVID Tracker dashboard (Figure 1) has been successfully built and demonstrated to key stakeholders. In addition to enabling spatio-temporal analysis and visualisation of public sentiment trends for key COVID-19 themes, the dashboard contains an innovative 'live' search feature that performs instantaneous topic and sentiment analysis for selected keywords over the past 7 days. Top 'positive' and 'negative' keywords and 'bags of words' (short phrases) are also extracted, with their sentiment determined in the context of the respective (positive/negative) social media post in which they occur.
- The HITL AI tool prototypes for COVID-19 topic and sentiment classification were validated using datasets from the SG Public Dialogue (on-line engagement exercise) held in May 2020. Successful validation demonstrated the prototype tools utility for practical deployment, to support rapid analysis of public comments in future SG public engagement exercises.

RAPID RESEARCH IN COVID-19 PROGRAMME

- Preliminary results showed UK public attitudes towards COVID-19 vaccination had been mostly positive (more so on Facebook as compared to Twitter) over the period of study: March 2020-November 2020. Public optimism over vaccine development, effectiveness and trials as well as concerns over safety, economic viability and corporation control were identified. We compared our findings to surveys (including YouGov, Ipsos MORI) and found them to correlate broadly.
- Our research showed that the proportion of messages expressing negative sentiment connected with the impact of COVID-19 on minority ethnic communities was higher on Twitter than Facebook, during the March 2020-October 2020 period studied. Sentiment trends were computed using our developed AI models, and found to correlate broadly with key events impacting ethnic minorities, such as the BAME report published by the UK Government, George Floyd and ethnic minority deaths.
- We analysed the suitability of AI-enabled social media analysis of Facebook and Twitter to understand public perceptions of COVID-19 contact tracing apps in the UK, over an eight-month period, from 1st March to 31st October 2020. Overall, the average positive sentiments were found to far outnumber the negative sentiments, with the majority of negative sentiments being in the North of England. These were found to broadly corroborate with findings from independent surveys that showed strong support for contact tracing apps. The sentiments varied over time, likely influenced by ongoing public debates around implementing app-based contact tracing using a centralised model where data would be shared with the health service, versus de-centralised contact-tracing technology.
- An initial pilot study was conducted on 'Long-Covid' topics that revealed a growing prevalence of social media conversation surrounding the issue. A preliminary manual search found dozens of support groups in Scotland and the UK on Facebook and Twitter, engaging tens of thousands of members. Analysis of long-COVID related keywords resulted in around 250,000 Facebook and Twitter posts from the UK, almost a quarter of which were from Scotland. The two most common symptoms were found to be 'fatigue' and 'shortness of breath'. 'Mental health' was most frequently associated with negative sentiment and concern, whilst 'support groups' was one of the most positively discussed topics. Further in-depth analysis of lived experiences is proposed.
- Iterative feedback solicited through demonstrations to focus groups, that included clinicians and members of the SG Communities Analysis Division and COVID-19 Modelling and Analysis Hub, helped refine the design and functionality of the prototype dashboard and AI tools. These are being utilised by SG analytical teams as a potentially valuable resource for policymakers, to help understand public mood and wellbeing during the pandemic, in near real-time.

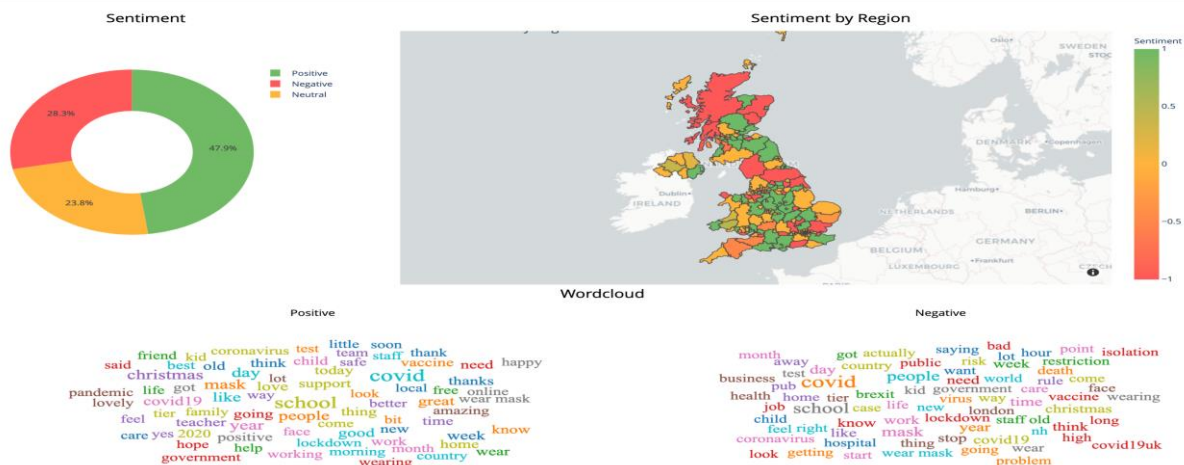


Figure 1: COVID Tracker dashboard illustration

RAPID RESEARCH IN COVID-19 PROGRAMME

WHAT IMPACT COULD THE FINDINGS HAVE?

Our findings will support decision makers by informing the development of appropriate interventions, such as public health education measures to dispel identified myths and misinformation. They could also inform development of exit strategies from current and future lockdowns, and support planning and decision making in other controversial public policy areas. These may help promote a learning public health policy cycle, where ideas can be tested against public attitudes before they are implemented, thereby maximizing their effectiveness and real-world applicability. Continuing use of our prototype tools by the SG has enabled near real-time assessment of public mood and well-being during the pandemic, and will support rapid analysis in future SG public engagement exercises.

HOW WILL THE OUTCOMES BE DISSEMINATED?

The outcomes are being presented in research articles in academic journals. Current list includes:

- [1] Hussain A, Sheikh A (2021) Opportunities for AI-enabled social media analysis of public attitudes towards Covid-19 vaccines, *NEJM Catalyst Innovations in Care Delivery*, DOI: [10.1056/CAT.20.0649](https://doi.org/10.1056/CAT.20.0649)
- [2] Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, Ali A, Sheikh A, (2021), Artificial intelligence-enabled analysis of UK and US public attitudes on Facebook and Twitter towards COVID-19 vaccinations, *Journal of Medical Internet Research*, DOI: [10.2196/26627](https://doi.org/10.2196/26627)
- [3] Cresswell K, Tahir A, Sheikh Z, Hussain Z, Hernández A D, Harrison E, Sheikh A, Hussain A, (2021), Artificial Intelligence-enabled analysis of social media data to understand public perceptions of COVID-19 contact tracing apps, *Journal of Medical Internet Research*, DOI: [10.2196/26618](https://doi.org/10.2196/26618)

We have published our collected datasets, which contain over 13 million Twitter messages and 5.6 million Facebook posts along with annotated results of our categorisation process. The datasets are publicly available via the ENU repository, and are being deposited in the [Health Data Research \(HDR\) UK Gateway](#) and HDR UK [BREATHE Hub](#). The dashboard is freely available to scientific communities and key stakeholders, and access to it can be requested through the [project website](#).

CONCLUSION

AI-enabled social media analysis techniques are currently underutilised in healthcare and yet to inform public policy research. Due to their current limitations, we argue these should be considered for adoption by institutions and governments alongside surveys and other widely used methods of assessing public attitude. Our project has afforded a timely opportunity for the SG to be the first in the world to successfully exploit prototype HITL AI tools that incorporate health and public policy expert knowledge to understand and inform public dialogue and engagement. The prototype tools are being used by SG analytical teams to support and automate near real-time understanding of online public discourse and sentiment at scale, and enable dynamic assessment of rapidly evolving SG policies in the current pandemic and beyond. We look forward to continuing working closely with the SG, ready to support future public engagement plans, providing further opportunities for the public, including underrepresented on-line groups, to get their voices heard.

RESEARCH TEAM & CONTACT

(Co-Leads) Prof. Amir Hussain, Edinburgh Napier University & Prof. Aziz Sheikh, University of Edinburgh	Email address: A.Hussain@napier.ac.uk aziz.sheikh@ed.ac.uk
Address: Edinburgh Napier University, EH10 5DT	Phone number: 0845 260 6040

ADDITIONAL INFORMATION

The project was funded with £135,104 by the Chief Scientist's Office and was successfully completed on the 17th of November 2020.