Knowledge Hub

18 June 2025

CENTRE FOR INDEPENDENT JOURNALISM

Advocating Media Freedom and Access to Information

Tracking Hate: Purpose, Data, and Lessons from CIJ's Social Media Monitoring Projects

Presenter

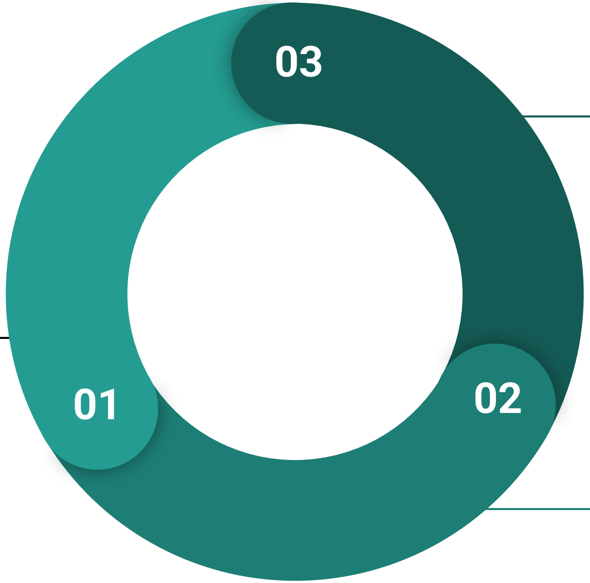
CENTRE FOR INDEPENDENT JOURNALISM

Advocating Media Freedom and Access to Information

Purpose

Legal Framework against hate speech

Addressing hate speech balances freedom of expression with protection from harm. International laws allow restrictions on speech inciting violence or discrimination



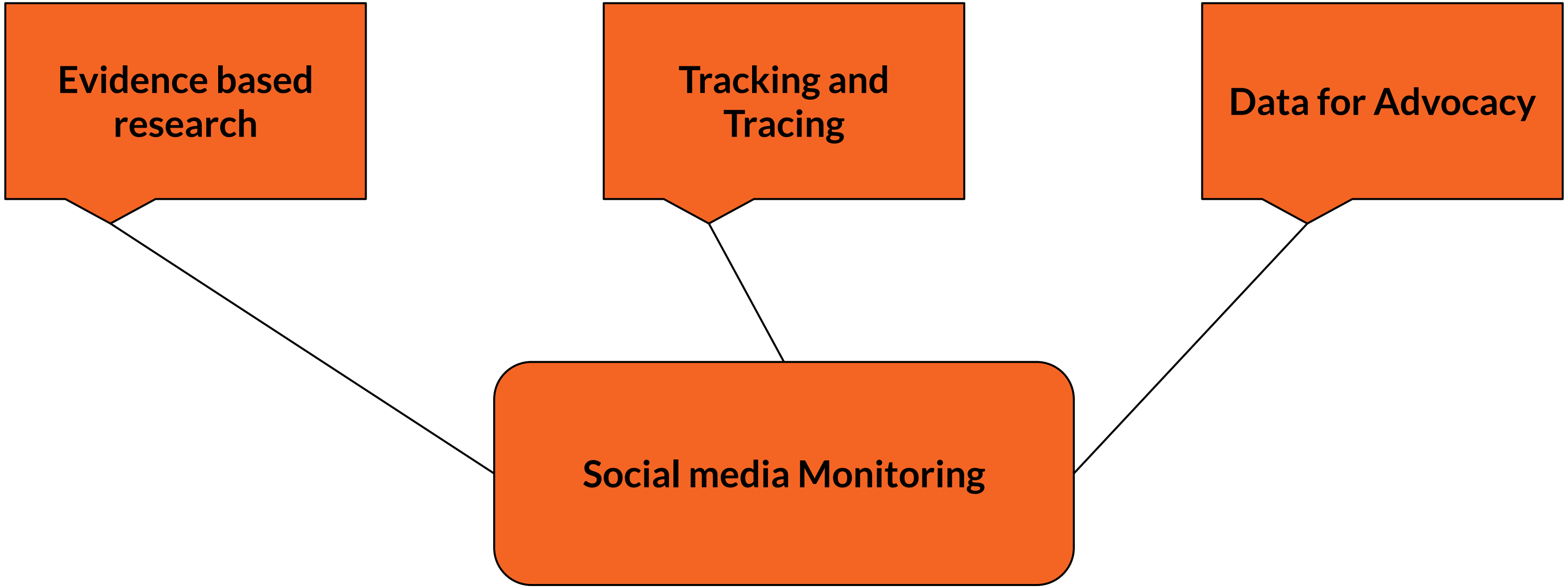
Spread of hate through social media

Platforms like Facebook, X, and TikTok historically exacerbates hate speech due its poor community standards, lackluster content moderation, engagement algorithms and unregulated monetisation systems.

Normalisation of Hate Speech

Hate Speech normalised through each interaction due to institutionalised and systemic racism - culture and narratives.

Purpose of data



Methodologies

- Robust Framework on how we categorise hate speech.
- Derived from grounding international standards to local context.

Disagreement/ non-offensive	Offensive/ Discriminatory	Dehumanising/ Hostile	Incitement/ call for violence
Different ideas, beliefs, opinions, generalisations,	Sexist Racist Negative stereotypes Slurs and vulgarism	Dehumanisation Sub-human comments	Attacks on minorities, people with specific characteristics Suggestions for bodily harm/ damage/ death
e.g. I don't like buying from non-Muslim traders.	e.g. Don't buy products sold by Cina because they are dirty	e.g. Don't buy products sold by *Cina babi* (pig) because they are "cancerous tumour" slowly eating up our economy	e.g. Go burn shops owned by *Cina babi* (pig)
e.g. I don't like migrants	Migrants are mostly criminals	Migrants are pests	"whack" the migrant children knocking on cars next time you see them.

CENTRE FOR INDEPENDENT JOURNALISM

Advocating Media Freedom and Access to Information

How

In 2021, our approach to online hate speech monitoring during General election was basic but structured. We relied on scraping data using keywords and character embeddings. Human monitors tagged and categorised posts, submitted reports, and flagged bot-like behaviour. Platforms covered included Facebook, Twitter.

By 2024, the methodology evolved into a layered, real-time monitoring system. We added new platforms like TikTok, expanded unit analysis, integrated image/video analysis, and started tracking AI-generated content.

CENTRE FOR INDEPENDENT JOURNALISM
Advocating Media Freedom and Access to Information

CENTRE FOR INDEPENDENT JOURNALISM
Advocating Media Freedom and Access to Information

Deployed Automated & Manual scrapping methodologies

Automated

- Contracted Zanroo as tool of choice for automated scrapping
- Content is keyword based
- No sentiment analysis due to limitations and nuances in the data.
- Evolved from Static keyword scraping to using dynamic keyword lexicons (update as we go system)
- We also went from Actor as first point on scraping to keyword.
- Content scraped were then verified through a robust internal system checking system

Manual/human monitoring

- Deployed an effective workflow and method to manual monitoring to make sure content is scraped as intended and to keep morale high
- Day-to-day, team leads distributed monitoring assignments using standardised templates, switching assignments every week in an effort to avoid fatigue and keep staff engaged. They further created weekly briefs of key trends, trending content, breaking narratives, and salient actors, which were distributed to the project coordinator. These helped inform weekly internal reports submitted to the project manager, allowing for timely strategic choice.
- Used the algorithm in our favour, through algorithm manipulation we were able to build a profile that suited our monitoring needs. (proved to be highly effective)
- Benefits of this methods is that we are able to look at content quicker and to see what is viral at the point of monitoring.
-

CENTRE FOR INDEPENDENT JOURNALISM
Advocating Media Freedom and Access to Information

How

Methodology – Evolution Over Time
Quantitative Improvements:

- Still use customised tools for scraping, but now with dynamic, evolving keyword lexicons.
- Introduced real-time alerts for faster response.
- Broadened platform coverage, including Gen-AI media content.

Qualitative Improvements:

- Moved from weekly checks to ongoing human-machine collaboration.
- Included deeper context analysis: sarcasm, irony, coded language.
- Monitors now assess multimedia content and cross-reference with bot behaviour and known narratives.

How

From Themes to Ecosystems

2021: We focused on individual posts related to hate and misinformation.

2024: We now examine the *ecosystem*:

- Track coordinated campaigns and super-spreader accounts.
- Monitor state and non-state actors.
- Analyse cross-platform content flow and amplification.
- Added layers: threat to social cohesion, national security, and economic impact.

How

New Areas of Focus

- **AI-Generated Content:** Detection and verification of synthetic media.
- **Network Analysis:** Mapping influence networks and echo chambers.
- **Community Input:** Integrating voices from the Rohingya and migrant communities.
- **Policy Relevance:** Tying online narratives to real-world policy impacts.

CENTRE FOR INDEPENDENT JOURNALISM
Advocating Media Freedom and Access to Information

How

2021	2024
Text/image based scrapping (majorly)	Multimedia + GenAI content
Static Keywords	Dynamic Keyword lexicon
Real-time alerts (Rapid Response)	Real-time alerts maintained
Actor focused	Network and Ecosystem analyis
Focus was on Facebook and Twitter	Includes TikTok
No impact tracking	Depending on issues theres analysis on Policy, Security, Economy and community.

Challenges

- Irrelevant content Scrapped
- Tool experiences downtime
- Mental Fatigue from human monitors
- Bot identification
- Missing out on harmful posts
- Reporting of posts - lengthy
- Resource intensive

CENTRE FOR INDEPENDENT JOURNALISM
Advocating Media Freedom and Access to Information

Lessons Learnt

- Social Media Analysis is tricky - getting on top of how the business model, Modes of engagement, Types of content and Actors is an ever evolving process.
- Never do it alone! Gather as many like minded partners to be part of your team. Leveraging different expertise in different fields is essential to a successful monitoring project.
- Having a well-thought out workflow is crucial to minimise mental fatigue, productivity and quality of content scraped.