



UPDATE ON THE INFORMATION DOMAIN

Issue 06/24 (June)

Hate Speech as a tool for Disinformation and Misinformation

INTRODUCTION

1. Hate speech, disinformation and misinformation rely on emotionally charged messages that evoke strong reactions from the public to cause chaos and confusion. The rapid dissemination of information over the internet and social media in the digital age has facilitated the spread of hate speech, disinformation and misinformation. Malicious actors make use of these as tools to undermine social cohesion, sow discord and manipulate public opinion.

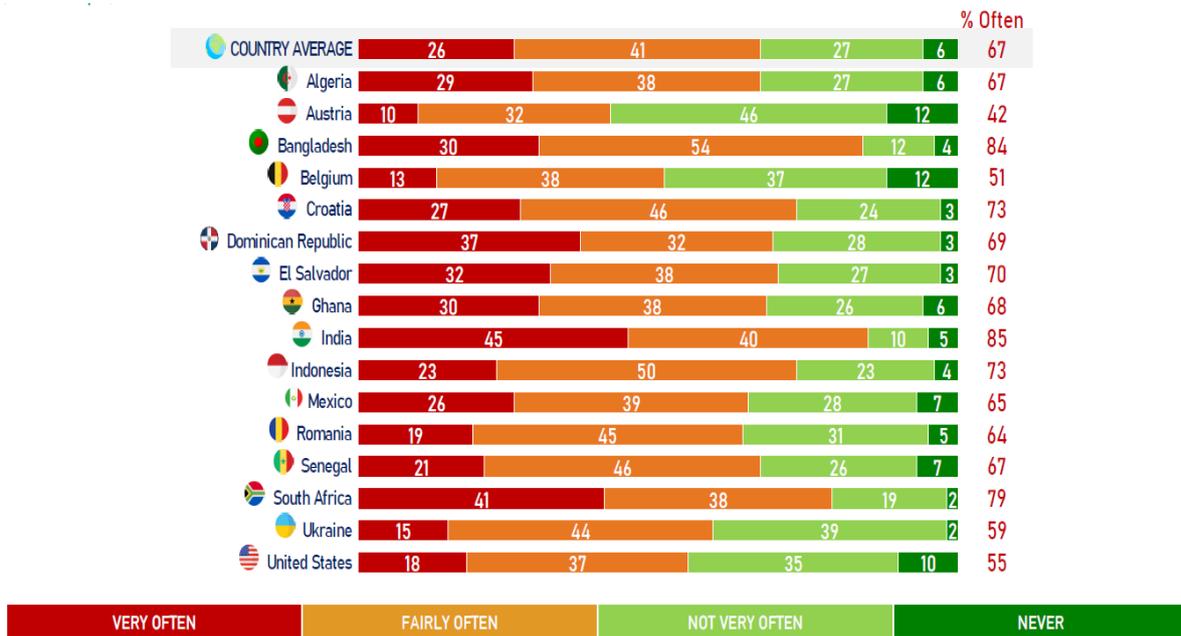
What is Hate Speech?

2. The *United Nations* defines hate speech as “any kind of communication in speech, writing or behaviour, that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other identity factor”. The boundaries of what constitutes hate speech is unclear and may vary according to different cultures and legal jurisdictions but its impact is universally detrimental.

Where is Hate Speech disseminated?

3. Social media is the main distribution platform of hate speech. According to a 2023 survey of 16 countries by Ipsos Group and UNESCO, 67% of internet users often come across online content that seems like hate speech (Figure 1).¹

Figure 1: Exposure of Internet Users to Hate Speech Online



Source: Ipsos-UNESCO Study on the impact of online disinformation during election campaigns, September 2023

4. Traditional forms of media such as radio, print and television are typically regulated and licensed to ensure content standards, technical quality, accuracy and fairness in the presentation of news and information. This includes an obligation to correct significant errors, so as to avoid misleading, deceptive and offensive content that incites violence, hatred or discrimination. Unlike traditional media where information typically flows one way from publishers to readers, social media platforms allow publishers and readers to interact and engage; this ‘to-ing’ and ‘fro-ing’ creates more avenues for hate

¹ Ipsos and UNESCO conducted a survey titled “A Global Survey on the Impact of Online Disinformation and Hate Speech”, published in September 2023. The 16 countries covered in the survey was chosen because a national election is foreseen in 2024.

speech to surface in the digital space. The use of legislation to regulate content on social media platforms with the intention of curbing online hate speech is challenging and controversial, as it is perceived by some to be an infringement on freedom of speech.

Why is Hate Speech dangerous?

5. Hate speech presents a real danger when physical actions are taken in response to it. Malicious actors can exploit the emotions and underlying fears among people, and leverage hate speech as a tool to propagate their misinformation and disinformation. Politicians may also use hate speech to drum up support for their causes, especially during sensitive periods such as elections.

6. For example, *Malaysiakini*² reported that multiple instances of hate speech were used during the campaign for Malaysia's 15th General Election in 2022. It reported that the leader of the Perikatan Nasional (PN), Muhyiddin Yassin had claimed that their political opponents, Pakatan Harapan, were "agents of Jews and Christians who were out to colonise Malaysia." *Malaysiakini* also reported that Muhyiddin claimed that "(i)n Kuala Lumpur and Petaling Jaya, there are Malays attending church on Sundays. They have become Christians." A video of this speech was shared widely on Tiktok ([Figure 2](#)).

² Malaysiakini is an online news portal from Malaysia established in 1999 and known for its independent, critical and controversial reporting.

Figure 2: Screenshot of Muhyiddin's Speech



Source: @phan.fizs on Tiktok

https://www.tiktok.com/@phan.fizs/video/7166771659779755290?referer_url=www.malaysiakini.com%2Fnews%2F644735&refer=embed&embed_source=%3Bnull%3Bembed_blank&referer_video_id=7166771659779755290 , reported by Malaysiakini, 18 November 2022

7. Similarly, *NBC* had, in an April 2024 report accused Indian Prime Minister Narendra Modi of using hate speech in his campaign in an attempt to secure votes. The report alleged that Prime Minister Modi had referred to Muslims as ‘infiltrators’ during the Indian General Elections and warned voters to not vote for the Indian National Congress as they will award Muslims “first right of the country’s resources”.

Use of Troll Factories

8. Malicious actors have used troll factories to boost their ability to generate hate speech online. They take advantage of large groups of automated bots to spread messages of hate to large audiences. These technologies are built with sophisticated algorithms and are able to provide greater visibility to online posts that potentially promote misinformation and disinformation. In Sweden, the Sweden Democrats' party and a conservative media group, Riks, were found to have jointly run a far-right troll factory, spread over multiple social media accounts on Facebook, YouTube, Tiktok and Instagram. *EU Disinfo Lab* reported that the content generated largely contributed to “sowing division and hate by impersonating Arab communities in Sweden.”

HOW TO TACKLE HATE SPEECH

Legislation as a Tool

9. Hate speech is a difficult issue to tackle, especially since legislation against it may infringe upon a citizen's rights to freedom of speech. Some countries have strong civil society and advocacy bodies, where such an anti-hate speech law will be seen as an attack on freedom of speech. However, legislations may be introduced to protect the social fabric and social harmony of a country so as to reduce the risk of hate speech. In Singapore, Senior Minister and Coordinating Minister for National Security, Teo Chee Hean was quoted by *Channel News Asia* in a 2022 report, stating that Singapore takes a “zero tolerance approach on hate speech and must ensure its communities are resilient against extremist rhetoric.” The Singapore Government's approach is to put in place different laws to safeguard the multi-racial and multi-religious fabric of its society. This includes the 1948 Sedition Act which criminalises acts and speech that promote hostility between different races or classes; Section 298A of the Penal Code, which addresses actions that promote enmity between different groups on grounds of religion, race or language; as well as the

Maintenance of Religious Harmony Act, which provides measures to prevent and manage religious conflict.

Education

10. Improving understanding on the dangers of hate speech and promoting multiculturalism and harmony serve to increase understanding between communities and reduce the effectiveness of hate speech-driven disinformation and misinformation campaigns. It is key to educate the public on internet norms and digital literacy to ensure that they are able to identify inflammatory content as well as to discern between legitimate and artificial online discourse. This contributes towards inculcating a zero tolerance for hate speech and report acts of hate speech. UNESCO published a guide in 2023 on *Addressing hate speech through Education* to encourage policy makers promote the usefulness of education policies against hate speech. For example, the Australian anti-racism education programmes rolled out in schools across the country are designed to teach the importance of countering racism, prejudice and discrimination from an early age. The guide also highlighted the case of Rwanda, where students are taught the horrors of the Rwandan genocide throughout their education journey, reinforcing the importance of preserving harmony and peace among different races. Such programmes aim to create a strong bedrock of multiracial and multireligious values among the youth.

Technology as a Tool against Hate Speech

11. Technology such as Artificial Intelligence (AI) which uses sophisticated large language and machine learning models, can be used to filter and detect social media posts with hate speech. This reduces the lag time in detecting hate speech driven disinformation and misinformation. The University of Waterloo has developed a machine learning tool, known as the Multi Modal Discussion Transformer (mDT), that has an 88% accuracy in detecting hate speech on social media platforms. This greatly reduces manual trawling of information and outperforms other language models that

may wrongly classify terms as forms of hate speech. Another example is BicepsHate – a deep learning model that is able to analyse and make sense of big data and linguistic context to classify hate speech.

Regional Dialogue to Raise Awareness on Hate Speech

12. Regional bodies such as ASEAN can help to raise awareness of hate speech spread through disinformation. This includes increasing awareness of the potential risks and threats and providing a platform to discuss potential concerns and solutions regarding anti-hate speech practices adopted by neighbouring countries. The ASEAN Intergovernmental Commission on Human Rights (AICHR) held multiple regional consultations on Freedom of Opinion and Expression in ASEAN. Some of the issues discussed include gendered disinformation, hate speech, cyber vulnerability, and digital transformation. Such dialogues provide opportunities to learn from past experiences in dealing with hate speech.

CONCLUSION

13. The development and promotion of trust building mechanisms domestically and regionally are essential for raising greater awareness of the dangers posed by hate speech and its use in disinformation and misinformation campaigns. States should also aim to stamp out hate speech by educating the public in internet norms, fostering friendly ties between groups of different faiths and backgrounds, to maintain harmony, peace and security in a multi-racial and multi-religious region.

.....

CONTACT DETAILS

All reports can be retrieved from our website at www.acice-asean.org/resource/.

For any queries and/or clarifications, please contact ACICE at ACICE@defence.gov.sg.

Prepared by:

ADMM Cybersecurity and Information Centre of Excellence

REFERENCES

News Articles

1. @phan.fizs on Tiktok
[Link: https://www.tiktok.com/@phan.fizs/video/7166771659779755290?referer_url=www.malaysiakini.com%2Fnews%2F644735&refer=embed&embed_source=%3Bnull%3Bembed_blank&referer_video_id=7166771659779755290]
2. AICHR hosts another regional consultation on freedom of opinion and expression in ASEAN
[Link: <https://aichr.org/news/aichr-hosts-another-regional-consultation-on-freedom-of-opinion-and-expression-in-asean/>]
3. Assessing hate speech through Education
[Link: https://unesdoc.unesco.org/in/documentViewer.xhtml?v=2.1.196&id=p::usmarcdef_000384872&file=/in/rest/annotationSVC/DownloadWatermarkedAttachment/attachment_import_26004d6b-ae0f-48cc-b394-06109eaaef51%3F_%3D384872eng.pdf&updateUrl=updateUrl7280&ark=/ark:/48223/pf0000384872/PDF/384872eng.pdf.multi&fullScreen=true&locale=en#1200_22_hate_speech.indd%3A.86131%3A658]
4. BiCapsHate: Attention to the Linguistic Context of Hate via Bidirectional Capsules and Hatebase
[Link: <https://ieeexplore.ieee.org/document/10022007>]
5. Disinfo Update 28/05/2024
[Link: <https://www.disinfo.eu/disinfo-update-28-05-2024/>]
6. En Suède, une enquête sur la fabrique à trolls de l'extrême droite provoque une crise dans la majorité
[Link: https://www.lemonde.fr/international/article/2024/05/14/en-suede-une-enquete-sur-la-fabrique-a-trolls-de-l-extreme-droite-provoque-une-crise-dans-la-majorite_6233243_3210.html]
7. Hate Speech, Mis- and Disinformation
[Link: https://www.un.org/sites/un2.un.org/files/notohate_fact_sheets_en.pdf]

8. Hate speech versus Freedom of Speech
[Link: <https://www.un.org/en/hate-speech/understanding-hate-speech/hate-speech-versus-freedom-of-speech>]
9. In Muar, Muhyiddin warns 'Jewish and Christian' agenda
[Link: <https://www.malaysiakini.com/news/644735>]
10. Modi accused of hate speech for calling Muslims 'infiltrators' at a rally days into India's election
[Link: <https://www.nbcnews.com/news/world/india-modi-accused-hate-speech-calling-muslims-infiltrators-rcna148916>]
11. Singapore's latest efforts in regulating online hate speech
[Link: https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=4879&context=sol_research]
12. Singapore takes zero tolerance approach on hate speech: Teo Chee Hean
[Link: <https://www.channelnewsasia.com/singapore/singapore-zero-tolerance-hate-speech-teo-chee-hean-2720241>]
13. Survey on the impact of Online Disinformation and Hate Speech
[Link: <https://www.ipsos.com/sites/default/files/ct/news/documents/2023-11/unesco-ipsos-online-disinformation-hate-speech.pdf>]
14. UNESCO dedicates the International Day of Education 2024 to countering hate speech
[Link: <https://www.unesco.org/en/articles/unesco-dedicates-international-day-education-2024-countering-hate-speech>]
15. What is Hate Speech?
[Link: <https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech>]