



Facebook in Myanmar: A Human Problem that AI Can't Solve

BY TEACIRCLEOXFORD | NOVEMBER 6, 2018

Mish Khan and **Sam Taylor** evaluate the utility of AI in detecting hate speech online.

Facebook's [inability](#) to effectively monitor and remove hate speech in [Myanmar](#) has come under fire by scholars, activists and media in light of the Rohingya genocide. These monitoring failings are partially a result of the network's heavy reliance on fallible Artificial Intelligence (AI) to conduct the otherwise more sophisticated work of human content moderators. This piece expands on the previous posts of [Ashley S. Kinseth](#) and [Francois Guillaume-Jaeck](#) on the topic of social media in Myanmar, through evaluating the current utility of AI in detecting hate speech.

Facebook's Major Local Presence and Abuse

While [Facebook policy](#) disallows hate speech, Facebook's overwhelming popularity in politically turbulent Myanmar means the platform is widely abused to disseminate hate speech and fake news – particularly anti-Muslim and anti-Rohingya sentiment—with disastrous consequences. The UN describes the mass exodus of over 700,000 Rohingya from Myanmar as a [“textbook example of ethnic cleansing”](#) and recently determined that top military officials be prosecuted for committing genocide, [referencing the role of Facebook](#) in propagating hatred towards the vulnerable group. A recent [Reuters investigation](#) found more than 1,000 examples of posts, comments and pornographic images attacking the Rohingya and other Muslims that had for years evaded the platform's content monitoring and remained on Facebook in Burmese language.

Facebook has been both reluctant and, more recently, incapable of addressing the platform's role as a medium for inciting violence in Myanmar. While the company is eager to enter new developing markets for profit, its systems are not equipped to handle complex language translation, to understand localised political knowledge, or to account for abysmal levels of media literacy in these locations. [For example](#), Facebook's auto-translation system has problems translating Burmese, providing bizarre results; one post referenced in the Reuters investigation was translated from Burmese saying “Kill all the kalars that you see in Myanmar; none of them should be left alive” into

English as “I shouldn’t have a rainbow in Myanmar.” Facebook **subsequently disabled** the Burmese auto-translation feature for ordinary Facebook users after the Reuters investigation was published.

Myanmar’s civil society groups **claim** that Facebook AI’s ability to proactively identify offending content is highly limited, given that merely tracing certain keywords such as “Ma Ba Tha” (a Buddhist group widely associated with religious nationalism) may not accurately reflect the content of posts. For example, the term ကုလား (**Kalar**) which is commonly used as a slur against Muslims, was banned and then un-banned in 2017, after Facebook realised it unintentionally censored several unrelated posts containing the term. The root word references the Indian subcontinent, and is sometimes also used to reference foreign qualities in general, and is used in many common words such as ကုလားပဲ kalar-pe (chickpeas), ကုလားထိုင် kalar-tain (armchair), or ကုလားတည် kalar-de (Indian style pickled fruits/vegetables). The dual usage of Zawgyi and Unicode encoding for Burmese text further makes auto-detection difficult. Rather than rendering all Burmese text through Unicode, the international standard for character encoding, Myanmar **uses a second system called Zawgyi** which is most popular on Facebook. It is thus not surprising that Facebook’s primary way of approaching hate speech has been an overreliance on third parties such as civil society organisations, who flag content for review to a small team of Burmese-speaking staff, none whom are based in Myanmar—in fact, Facebook has no presence within Myanmar full stop.

AI’s Fundamental Limitations in Addressing Hate Speech

Even in English alone, prior to translation, natural language processing (NLP)— that is, the area of computer science attempting to mimic human linguistic interpretation— comes up against multiple barriers to successful detection and struggles to find sufficient context to make well-informed content removal decisions. Keywords can form the basis of such contextual analysis, but often stand separate from the surrounding post without NLP necessarily having the capability to coalesce it in any meaningful way. An accompanying sentence or paragraph to a keyword may not even be sufficient, with real knowledge of real world events also needed for a considered determination to be made. **Throw in a word like “love”**, and the system might dismiss the entire flag, by automatically deeming the sentiment to be positive through such a red herring.

Complementing these context processing issues, a recent New Scientist investigation reported that Google’s AI detector could be foiled simply by adding typos or spaces, eliminating even effective keyword-only identification. Perpetrators could easily seize upon such apparent oversights and outsmart systems, leaving a commitment to employing well trained and vigilant human content moderators as the only measure Facebook could implement with reasonable prospect of success. Setting parameters too broadly leaves open the potential for over policing, the removal of innocuous content, or worse still, censoring peace promoting sources instead of spiteful messaging.

And this is just on a textual basis, not accounting for graphic photo and video content which also harbours hate speech. [Rosetta](#) AI was recently introduced and touted as Facebook's solution on this front by being able to extract text from images on whatever surface it is printed. However, it does not explicitly address the [underlying issue](#) that the actions depicted in those same images or clips may themselves be the most flagrant incidents of discriminatory displays, and these fall outside of Rosetta's stated capabilities. Whatever text is extracted is also hence subject to the same context constraints NLP and machine learning models currently face.

Can External Reporting Help?

Even with the overreliance on external reporting designed to combat the lack of tangible local monitoring, reporting rates in Myanmar have remained low as Facebook's [reporting tools were not translated to Burmese until May this year](#), nor was the ability to report content circulated through Facebook messenger (also rolled out [in May](#)).

After years of inaction, during which Facebook devoted scant resources to Myanmar despite multiple warnings of potential problems, Facebook is now starting to take its role in minimising the abuse of its platform more seriously. While in early 2015, Facebook had employed just four Burmese-speaking staff worldwide, it now employs about [60 overall](#) and wants to employ [100](#) by the end of the year—although these Burmese-speaking staff remain entirely based outside of Myanmar, with around 60 part of a secretive operation in Kuala Lumpur codenamed "[Project Honey Badger](#)". Furthermore, after an [internal investigation](#) in August, Facebook banned 20 individuals and organizations from Facebook in Myanmar — including Senior General Min Aung Hlaing, commander-in-chief of the armed forces, and the military's Myawaddy news network, stating: "We continue to work to prevent the misuse of Facebook in Myanmar... This is a huge responsibility given so many people there rely on Facebook for information — more so than in almost any other country given the nascent state of the news media and the recent rapid adoption of mobile phones."

A Human Problem for Human Oversight

The above mentioned failings of Facebook's AI in detecting critical cultural nuances on its platform should make it clear to the company that it cannot outsource what is an intricately complex and indeed humanly nuanced problem. That's not to say well-constructed AI models shouldn't be pursued as a way of reducing human workload. Yet, only when AI systems contain a kind of realisable situational awareness can Facebook begin to delegate responsibility away from human monitoring. At the very least, it should flag content for human review rather than letting the algorithm resolve any offending content.

Until such a time AI can be programmed to contain broader conceptual understanding of content, and not mere computational analysis, Facebook should stand firm on its promises to prioritise

Burmese content moderator education. This measure ensures immediate gains and improves identification accuracy in the fight against hate speech. Properly training the algorithms of its AI systems, which are presently an expedient shortcut, can only be achieved in the long run and should not be attempted without fully realised technology.

Not only is this desirable in terms of devoting human resources to the countries it operates in, but it would also enable the potential development of community standards as to what is acceptable, fostered by collaboration between user-based reporting and designated review officers who know far more of local subtleties and underlying, innately human, motivations. While such standards themselves and sensible consensus around them will often be moot, especially given the divided situation in Myanmar, it is preferable to formulate them with local stakeholders at the centre of the dialogue, rather than leaving their determination as the product of machine based decision making.

[Image credit: [Flickr](#)]

Mish Khan is from the Australian National University College of Asia Pacific and is currently based at Yangon University under an Australian Government scholarship.

Sam Taylor is a legal researcher with an interest in regulation around AI ethics and algorithmic fairness. He is currently completing his studies in Law and International Relations at the Australian National University.

POSTED IN ESSAY • TAGGED AI, FACEBOOK, HATE SPEECH, MYANMAR, TRANSLATION

PREV

Tea Circle in Toronto: Shifting Southeast Asia(s)

NEXT

Critical Thinking in Myanmar's Education System (Part I)

SERIES: COVID-19 AND MYANMAR

ကိုဗစ်-၁၉ နှင့် မြန်မာရာသီဥတုပြောင်းလဲခြင်း
ဆိုင်ရာ လူငယ်တို့၏ တက်ကြွ လှုပ်ရှားမှု

COVID-19 Policy Response Needs and
Opportunities

Wav
r

TEA CIRCLE'S BOOK REVIEWS

Perspectives on War, Peace, and Rebel
Politics: Introduction

Myanmar's Buddhist-Muslim Crisis:
Rohingya, Arakanese, and Burmese
Narratives of Siege and Fear, by John Holt.
Honolulu, Hawaii. University of Hawaii Press,
2019. 301pp.

My
C
F
Gay

RECENT POSTS



Resilience to Dual Crises: The Coup And the Pandemic

March 30, 2022



Chronicle of a Coup: June 1, 2021

March 26, 2022

Chronicle of a Coup: May 28 & 30, 2021

March 18, 2022



TAGS

Book Review **burma** coup democracy education ethnicity Karen military military coup **myanmar** NLD
politics Rohingya transition Yangon

ARCHIVES

Select Month

▼

March 2022

M	T	W	T	F	S	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

« Feb

CONTACT US

 Center for Southeast Asian Studies
Asian Institute
1 Devonshire Place
Toronto, Ontario, M5S 3K7
Canada

 (+1) 416-946-8996

TEA CIRCLE ON FACEBOOK

FOLLOW BLOG VIA EMAIL

Enter your email address to follow this blog and receive notifications of new posts by email.

FOLLOW

FOLLOW ME ON TWITTER



Tea Circle

@TeaCircleOxford

The next instalment of [#ChronicleofaCoupe](#) is out:

Christopher J. Walker describes one attempt to help relieve the dire food shortage resulting from military rule in Myanmar. [#WhatsHappeningInMyanmar](#)

Read it here: tinyurl.com/juuesedt



[Embed](#)

[View on Twitter](#)

COPYRIGHT TEA CIRCLE 2018. ALL RIGHTS RESERVED.

The opinions expressed on this website belong to the authors alone, and do not reflect the views of the editors, the Centre for Southeast Asian Studies, the Asian Institute, the Munk School of Global Affairs and Public Policy or the University of Toronto.

Powered by WordPress.com.