



Social Media, Democracy, and Violence in South Asia: A Background Note

By Dr. Sanjana Hattotuwa, Special Advisor, ICT4Peace Foundation

Background note by Dr. Sanjana Hattotuwa for a panel presentation at an online seminar organised by the **Center for the Study of Organized Hate (CSOH)** on 30 September 2025 titled '**Big Tech and the Misinformation Crisis in South Asia**'.ⁱ

My research on the weaponisation and instrumentalisation of Facebook began in 2012, with initial studies on platform instrumentalisation by Buddhist violent extremism emerging in 2013. This work has tracked, for longer than most in the region, and beyond, how Big Tech platforms have evolved into industrial engines of systematic, structural violence across South Asia, where algorithmic amplification transforms disinformation into mass atrocities with devastating human consequences, whilst simultaneously exploiting and deepening pre-existing sociopolitical faultlines that these societies have struggled with for decades. This is occurring at the same time as social media powering key sectors of fragile economies, contributing to education, and more recently, upending workforce skills, and industry through AI. Defined by, *inter alia*, democratic deficits (ironically in a region with the oldest democracy in Asia – Sri Lanka), autocratic fiat, institutional capture, regulatory overreach, unfit-for-purpose laws, socio-political instability, sectarianism, various forms of violent extremism, cross-border, and internecine conflict, social media's location in polity, and society is complicated, and fluid – a dynamic that cannot be understood through a Western gaze that sees social media as the root cause of violence.

The enduringly, and intrinsically dual nature of social media platforms presents a fundamental paradox.

Discussions must move beyond requests for simple examples on social media harms, as the moderator asked of the speakers, to a more grounded, robust examination of the context around how social media's devastating impact on democracy exists simultaneously with how these platforms strengthen democracy and fight authoritarianism. To acknowledge this enduring duality is fundamentally important.

The same platforms used by autocrats and hate entrepreneurs serve, coterminously, rights activists and investigative journalists. Social media enabled civic resistance in Myanmar, the 2014 activism in Sri Lanka around a very consequential presidential election, the 2022 *aragalaya* movement, and more recent social mobilisations in Bangladesh, and Nepal. All these movements faced internet shutdowns or blocks, revealing how platforms have become critical infrastructure for both harm, in the hands of autocrats, and democratic participation, as a vector for a repressed *vox populi*.

This is not to say social media companies should escape scrutiny or accountability. Facebook's recommendation systems fuelled anti-Rohingya content that contributed to the 2017 Myanmar genocide, forcing 700,000 refugees to flee, whilst allocating just two Burmese-speaking moderators for over 20 million users. The platform leveraged longstanding ethnic tensions between the Buddhist majority and Muslim minorities that date back generations. Examined in my doctoral research, in early 2018, Facebook contributed to the worst anti-Muslim riots in Sri Lanka's history by amplifying existing Sinhala Buddhist nationalist grievances against Tamil and Muslim communities rooted in postcolonial ethnic conflict. WhatsApp enabled systematic mob violence across India, where at least 27 people died in lynchings between 2017 and 2020 after false child abduction rumours spread through forwarded messages, exploiting caste hierarchies, communal divides between Hindus and Muslims, and urban-rural tensions that platforms did not create but systematically weaponised through viral amplification. In April 2019, the post-Easter Sunday terrorism media, and information ecologies in

Sri Lanka (also studied for doctoral thesis) - where Meta's products, and platforms played a central role - also showed a similar narrative pathology, where incendiary content, and commentary spread without any discernible platform friction or oversight, and instigated hate, and harm towards all Muslims across the country. Twitter/X amplified AI-generated deepfakes during the May 2025 India-Pakistan conflict, where fabricated videos of Pakistani officials garnered over 750,000 views and mainstream media coverage, demonstrating how synthetic content can push nuclear powers towards catastrophic miscalculation by inflaming partition-era animosities and territorial disputes that have defined South Asian geopolitics for 75 years.

Four critical asymmetries make South Asia uniquely vulnerable to platform manipulation.

Resource discrimination means that 75 percent of internet users live in the non-English speaking Global South yet receive minimal content moderation resources, a situation worsened by Zuckerberg's early 2025 policy pivot and shift to AI oversight. The region has the highest number of internet and social media users in the world but must contend with the lowest allocation of human moderators and content review resources. **Cultural context blindness** describes how algorithms fail to detect Buddhist nationalist content openly promoting eliminationist rhetoric and violence through memes that present funny frames. **Disinformation targeting public health** sees malign actors mislead and misinform tens of millions, leading to violence, suffering, death, and intergenerational harms around women's health issues and vaccines. **Monopolistic vulnerabilities** arise where social media functions as "the internet" for first-time mobile users, aided by telecommunications deals around unlimited YouTube, Facebook, WhatsApp, Instagram, and TikTok access.

Two additional features of truth decay require attention.

Synthetic and manipulated media extend beyond deepfakes to encompass a range of fabricated content that exploits low verification capacity.

Influence operations by China add a foreign dimension to domestic manipulation, creating complex attribution challenges for content that inflames existing tensions.

The same technological architectures and network effects that amplify hate speech and misinformation enable prosocial and pro-democracy movements that authoritarian governments seek to suppress. Facebook, WhatsApp, and X have facilitated anti-corruption protests, human rights documentation, emergency disaster response coordination, and democratic mobilisation against military coups and authoritarian overreach. This creates a regulatory paradox where measures to restrict harmful content risk empowering state censorship against legitimate dissent. The so-called 'Arab Spring' demonstrated how social media could catalyse democratic uprisings, whilst Myanmar's military junta's internet shutdowns during the 2021 coup revealed how platforms serve as critical infrastructure for civil resistance. As noted earlier, Sri Lankans (and not just activists) used social media at an unprecedented scope, and scale to seed, and spread compelling critical narratives during the 2022 *aragalaya* protests that quicker than most expected, toppled the murderous, and corrupt Rajapaksa government. There are innumerable other examples from South Asia, and in just the past decade which complicate simplistic recommendations to introduce regulations, and laws to address online harms. Sri Lanka's draconian Online Safety Act (OSA) is a cautionary tale in this regard, posing a historically unprecedented threat to democracy, but at the same time, diverting attention away from what is an enduringly important challenge to meaningfully address the instigation of hate, and harms online – growing at pace.

Ever-increasing consumption coupled with low media literacy and democratic deficits creates exploitable vectors for domestic and foreign malign actors. The adoption and adaptation of Trumpian playbook tactics around online regulations and instrumentalisation of social media has spread across the region. South Asian leaders prefigured Trump's manipulation of information landscapes, but he now serves as a convenient excuse for authoritarian overreach that predates his influence.

Five interventions must address the dual-use nature of social media.

Rigorous, grounded academic studies on the interplay between offline autocracy and online harms must examine the role, reach, and relevance of social media platforms in specific regional contexts. **Algorithmic accountability** demands mandatory disclosure and economic disincentives for violent content whilst protecting political speech, creating transparency around recommendation systems that amplify harmful material. **Cultural and contextual grounding** requires local language moderators, Global Majority aligned AI systems, and regional partnerships that understand both harmful exploitation of divisions and legitimate grievance expression. **Rights-respecting regulation through co-regulatory approaches** must prevent both corporate negligence and state overreach, including reparative or restorative justice mechanisms for communities harmed by platform negligence. **Digital resilience building through literacy education, inoculation against disinformation, and political leadership** strengthens citizens' capacity to identify manipulation whilst preserving their ability to organise collectively. Interface dialogue programmes can help societies address underlying tensions without enabling censorship.

With nuclear-armed antagonistic neighbours, significant democratic deficits, and the world's largest population of young internet users, coordinated action across technological, regulatory, educational, and diplomatic dimensions must replace voluntary corporate initiatives. These platforms have become indispensable infrastructure for both harm and democratic participation, requiring nuanced governance frameworks that can distinguish between content that inflames pre-existing faultlines towards violence and content that challenges entrenched power structures through legitimate protest.

ⁱ Due to technical glitches on the end of the organisers, Sanjana Hattotuwa couldn't make it to the Zoom call.