ICT for Peace Foundation
Policy Paper

Christchurch Call
Paris, 14 May 2019

Through our work at ICT4Peace over the past years, we have witnessed and analyzed the changing use of social media and its growing impact on critical issues related to democracy, political stability, freedom, communication and security. The euphoria about the role of social media as a primarily positive force during the Arab Spring has given way to a much more layered and complex picture of its role and uses across society and around the globe. The sheer enormity of today's social media platforms, the volume of users and almost infinite mass of content, means that the containment of the spread of violent content, as witnessed after the Christchurch attack, proved almost impossible.

We have been working on these issues for many years now, including, launching on behalf of the UN Security Council the Tech against Terrorism platform with inter alia Facebook, Microsoft, Twitter and Telefonica; carrying out cybersecurity policy and diplomacy capacity building for inter alia ASEAN and the CLMV countries; working with the UN GGE and ASEAN on norms of responsible state behaviour for cybersecurity with the ASEAN regional Forum on CBMs; carrying out workshops in Myanmar, Sri Lanka on online content verification and online security; participating in the CCW GGE discussions in Geneva on Lethal Autonomous Weapons Systems (LAWS), and analyzing the role of artificial intelligence and its role in peace-time threats such as surveillance, data privacy, fake news, justice, the changing parameters of health including the risks of certain biotechnological advances and other emerging technologies.

The challenge of controlling and removing terrorist content online

Despite now serious attempts by social media platforms to control content that violates norms, human beings are simply unable to keep up with the speed and connectivity of content creation around the world. This task can only be computationally managed by algorithms and AI, but these are also opaque, offering biased recommendations, search functions and are in part responsible themselves for the rise in extremism, conspiracy theories and destabilizing content online. However there is some hope going forward in the engineering of greater friction in times of crisis. Instead of on/off censorship, engineering greater friction into sharing can help,

at scale, control and curtail the flows of misinformation. The best example of this comes from India and WhatsApp - https://www.reuters.com/article/us-whatsapp-india-fakenews/whatsapp-curbs-message-forwarding-in-bid-to-deter-india-lynch-mobs-idUSKBN1KA07V. For years, apps – linked to 'growth hacking' made it as easy as possible to share and engage with content. In countries and contexts with little to no media literacy, this because quickly weaponised by actors who used the virality of content over social media, without any fact checking whatsoever, as a vector for misinformation spread at scale. With added friction – limits on the number of forwards, adding visual aids, adding an extra step to interact, in the backend, through algorithmic suppression of content with poor quality (e.g. clickbait articles) are a range of ways that from the app (front facing) to the algorithm (back-end) companies can and have invested in ways that in effect, reduce mindless sharing. https://twitter.com/sanjanah/status/1127957760266645504.

Protecting democracy, ethical principles and a free open Internet

The challenge of balancing the need to maintain a free and open Internet with the need for security and protection of human beings, data, ethical principles, human rights and democratic processes is daunting. It is essential to achieve a broad alliance pushing for practical change to prevent the spread of extremist content online and the glorification of the perpetrators of mass murder. It is also important to protect users from fake news, misinformation and manipulation in particular in the terrorist context. From the Global South perspective, a key concern is also how and if measures undertaken by social media companies in response to Western demands, challenges and concerns, might undermine the ability of civil society to hold authoritarian governments accountable, and could weaponise processes and structures to clamp down on dissent. We must remain vigilant to ensure that at each step of the way these concerns are considered.

Social Media, Coming of Age

We have a responsibility to current and future generations to ensure a framework for all emerging technologies that respects basic human rights and ethical principles. Social media is evolving and society needs to figure out how these tools should be used now and in the future. There is a real risk of the race to the bottom with hate and the baser nature of human beings taking the lead as users gravitate toward clickbait and gruesome content. However, society seems to have fortunately reached an ethical border with the livestreaming of murder, just as we had reached a border in biotechnology with the cloning of a human being. We need to develop guidelines for how we, as a global society want to move and operate in the social media space. What kind of ethical principles need to be built into the algorithms and AI that will control our future content and interaction online? How should social media companies evolve in their approaches and business models to take into account the human dimension? We need to shift and develop technical measures that consider the quality, not just quantity, of conversations online, the mental and physical health and age of consumers and the ways in which content is shared and manipulated.

For years, Facebook and other Silicon Valley companies prioritised 'growth hacking' by which they meant that the increase in user base for products and platforms overtook every other aspect of the business. At Facebook this was led by Sheryl Sandberg, the company's COO. This is why the company is facing the issues it is today, and why Mark Zuckerberg's pivot to the health of conversation and content is significant. It is a major change of course for the company, including the new emphasis on privacy over sharing, a concept scoffed at by Zuckerberg himself in the past. Though the contours of what the company will become are clear, it is unclear what exactly will be done to ensure the quality of conversations and content are engineered to be biased away from the toxic, violent and hateful. But across major platforms including YouTube, Twitter and Facebook's Instagram, Messenger and WhatsApp platforms, there is a new emphasis on securing user privacy and at the same time engineering ways that also protect users from hate, harm and violence.

Added to this is a new interest, at the operating system level of phones and tablets on both iOS and Android, ways through which Apple and Google respectively are now keen to ensure users have a good balance of on-screen and off-screen time. This includes logging device and app usage at the OS level, and by engineering tweaks on apps like Instagram for example (first, deeply unpopular within Facebook) to give a visual indication of when new content was over, and the user was scrolling through content already seen or engaged with, thus ensuring the user spent less time on the app, not more. Less time on apps meant less advertisements seen, and less interactions with the app or operating system, in turn resulting in less monetisable action points, impacting, at scale, profits as well as the harvesting of user level engagement metrics. So, what appears to be a simple change is actually a major shift for companies that now prioritise the health of users over constant engagement and addiction to apps.

Proposed actions for Social Media companies and other key actors:

1. Ensure that all steps taken to control terrorist content do not undermine the ability of civil society to hold authoritarian governments accountable, or weaponise processes and structures to clamp down on dissent. This is key for any meaningful reform in this domain (where human security, user health and privacy is pegged to Silicon Valley's vaulting profit mode), to move from what is presently an antagonistic relationship to a more mutually beneficial one.

2. Develop joint multi-stakeholder taskforces to consider the big picture of the human dimension of social media and develop ethical principles that could guide Social Media companies and users in the online world.

3. Social Media companies in cooperation with government, law enforcement and civil society need to reinforce joint SWAT team responses for content that meets certain extreme criteria, e.g. in particular livestreaming of murder or other heinous acts.

4. Prioritize the development of AI that could better define and distinguish types of content and support in the clamping down in emergency situations on the connectivity of terrorist content. *(The real problem at the moment is the definition of terrorist content. No AI at present can easily distinguish between the media's coverage of a terrorist incident which may include graphic violence, and a terrorist group's promotion of violent ideology, which may include the same or similar graphic violence. AI can also be fooled, and without human review, can and has led to instances where content documenting human rights abuses in war have been entirely deleted, in effect contributing to the impunity of perpetrators.)*

5. Ensure existing AI and algorithms do not promote terrorist content and extremist views by pushing more such content to users. *(YouTube's recommendation engine / algorithm, widely and increasingly criticized for the promotion over time of increasingly extremist content, is undergoing major overhauls on these lines).*

6. Proactively remove hate speech and repeat offenders. *(This is directly linked to how much of human and technical resources can be put in by these companies, and also the challenge of end-to-end encrypted channels / platforms like WhatsApp, where even the company doesn't easily know the kind of content exchanged in groups.)*

7. Improve review mechanisms and responsiveness.

8. Reinforce trusted reporting network that expedites the flagging of content vetted through experienced individuals and institutions.

9. Improve the reporting mechanisms built into Facebook apps like Facebook Messenger to make it easier and simpler to report violent or hateful content.

"It is important in these difficult times also to remember that just as social media helps extremist ideology take seed and grow, it also helps in healing, empathy, gestures of solidarity, expressions of unity, the design of conciliatory measures and the articulation of grief and sympathy. In the immediate aftermath of the Christchurch attack, a cursory top-level study of the nearly 85,000 tweets generated in the 24 hours after the violence shows a global community outraged or dismayed at terrorism, an outpouring of love, empathy and solidarity, engagement that spans many continents and languages, addressing prominent politicians and journalists, featuring hundreds of smaller communities anchored to individuals based in New Zealand and in a manner overwhelmingly supportive of the Muslim community." Sanjana Hattotuwa, http://www.scoop.co.nz/stories/HL1903/S00124/pulse-points.htm