TECHNOLOGY

FIGHTING FOR DATA DREGS – AND LOSING THE FIGHT AGAINST DIGITAL VIOLENCE

Thursday, May 26, 2022

With more countries adopting legislation to tackle digital violence, there is hope in the battle against online abuse, write Sun Sun Lim of the Singapore University of Technology and Design and Roland Bouffanais of the University of Ottawa. With limited access to data on online content and behavior that could point to remedies, however, academia, civil society and the public sector can only do so much to improve the safety and wellbeing of internet users.

The attacker: Trolls can intimidate and threaten their victims on every available online platform through different communication modes and content

In the Asia-Pacific region, momentum has been building around legislation to tackle digital violence and other forms of online harm. In January 2022, Australia's Online Safety Act came into force, introducing mechanisms to remove harmful content, protect adults from online abuse, and augment existing protections for children against cyber bullying. Meanwhile, the Singapore government is set to launch legally binding codes of practice for technology platforms to establish government is set to launch legally binding codes of practice for technology platforms to establish accessible systems that allow users to report harmful content, take necessary remedial action, and regularly publish reports on the efficacy of their measures. In April this year, the Indonesian government passed the long-awaited Sexual Violence Bill, which recognizes nine forms of sexual violence as punishable acts, including physical and nonphysical sexual harassment, forced marriage, and, notably, cyber sexual harassment. Beyond these examples, other Asian nations are considering similar legislation.

Online harm and digital violence can assume many forms, have negative long-term effects on victims, and are rapidly becoming a societal scourge. Perpetrators can intimidate and threaten their victims on every available online platform through different communication modes and content. Trolls can send victims sexually explicit images on direct messaging platforms such as WhatsApp and Telegram. Participants in discussion forums such as Reddit can disseminate misogynistic and sexist memes disguised as dark humor. Gamers can perpetrate acts of sexual aggression against vulnerable players, including making unwanted sexual advances, rape jokes, or mounting virtual assaults in games such as *World of Warcraft* and *Rape Day*. Even the emerging metaverse has been acknowledged to have a "groping problem". The possibilities are endless, but so is the trauma – victims can suffer from mental illness, reputational damage, fear for personal safety and reluctance to go online, thus constraining their freedom to enjoy online interaction.

With so many life-changing innovations born every day, one would think that online harm and digital violence can be technologically resolved — or at least managed — through bots to detect adverse content, verification systems to ban bad actors, automated prompts to caution against aggression, and so on. But how close are we, truly, to achieving a desirable level of online safety through technological solutions?

The victim: In the Asia-Pacific region, momentum has been building around legislation to tackle digital violence and other forms of online harm (Credit: SB Arts Media / Shutterstock.com)

Given the scale and nature of the problem, the current technological approaches to dealing with online harm depend heavily on artificial intelligence (AI) and big data. To cope with the deluge of harmful content, technology platforms deploy machine learning to trawl through mountains of data to detect trends for building analytical models that are applied to image and speech recognition.

Facebook, for example, uses AI photo matching technology to detect the sexual exploitation of children. The potency and effectiveness of these AI-driven content monitoring and moderation techniques have not been proven, however. After all, they still belong to the state-of-the-art realm of machine learning. In practice, they must be ceaselessly refined, updated, and upgraded in the face of new ways of harming users, novel genres and emerging viral trends. The fight against online abuse led by Big Tech corporations is ultimately a Sisyphean endeavor.

Indeed, given the multi-faceted complexities of the roots and effects of online harm and digital violence, they are best tackled with the combined efforts of technology companies, enforcement agencies, academic researchers and civil society organizations. Currently, though, only technology companies have access to troves of data on online content and behavior that would illuminate possible remedies. The egregious problem of digital violence and online harm thus throws into sharp relief the need for greater data sharing and the grievous effects of data asymmetry among technology companies, academia, and civil society.

Technology behemoths such as Alibaba, Google, Meta and TikTok marshal considerable resources to undertake “surveillance capitalism” to distil from our “behavioral surplus” highly detailed, identifiable data capturing human activity, mobility, physiology, emotions, and sentiments in staggering detail. Governments, universities and think tanks can in turn only undertake modest, piecemeal data collection efforts that pale in scale, scope, duration and resolution to commercial endeavors. If data were the new oil, technology companies drink from the fountain of premium grade oil, leaving non-profit organizations to contend for oil dregs.

The crime scene: The secrecy offered by end-to-end encryption over messaging platforms has been exploited for illicit activities such as terrorism, drug trafficking and child pornography (Credit: Michele Ursi / Shutterstock.com)

To address the growing problem of online dangers, we need greater data sharing for enhanced monitoring, effective detection, improved enforcement, comprehensive victim support, and more encompassing third-party research into online platforms and their risks. But these all hinge on how forthcoming technology companies actually are with the vats of data they own and control. Tremendous complexities also arise around how data can be shared without compromising individual privacy. Worryingly, recent reports reveal that technology companies have fallen for fraudulent legal demands and shared user data that were subsequently used to harass and extort young victims. While these are difficult issues regulatory bodies must strive to iron out, they should not be used by data owners as an expedient justification to stonewall data-sharing imperatives.

Combating digital violence and online harm reveals another confounding factor: the secrecy promised by messaging platforms that run on sophisticated end-to-end encryption technology. Whereas secrecy can be a virtue, it presents a stumbling block to enforcement against online harm. When online harm occurs, victims must gather evidence. This also necessitates that enforcement agencies request relevant evidence from technology platforms. The growth of end-to-end encryption over platforms such as WhatsApp and Telegram, however, means that such requests are difficult to entertain. Indeed, the secrecy offered by such platforms has been exploited for illicit activities such as terrorism, drug trafficking and child pornography.

Digital violence is abhorrent, and with so much of our lives now lived online, it is vital that our online worlds are as secure and welcoming as we would like our physical worlds to be. In our increasingly hybrid online-offline existence, why should one world be less safe than the other? How can we trust the gatekeepers of our digital existences to care for our safety when their profits weigh heavily on algorithms that maximize user engagement and ad-selling at all costs? This wave of legislative action across Asia to address online harm is heartening, but as long as academia, civil society and the public sector are forced to make do with “data dregs”, very little headway can be made to improve the safety of platforms and the wellbeing of their users.

Further reading:

Henry, Nicola; and Powell, Anastasia. (June 16, 2016) “Technology-Facilitated Sexual Violence: A Literature Review of Empirical Research”, *Trauma, Violence, & Abuse,* vol. 19, no. 2, pp 195-208, SAGE Publications, Thousand Oaks, CA, USA.

Lim, Sun Sun; and Bouffanais, Roland. (January 25, 2022) "'Data dregs’ and its implications for AI ethics: Revelations from the pandemic”, *AI and Ethics,* Springer Nature, Cham, Switzerland.

Vickery, Jacqueline Ryan; and Everbach, Tracy (eds). (2018) *Mediating Misogyny: Gender, Technology, and Harassment,* Palgrave Macmillan, London, UK.

Zuboff, Shoshana. (January 31, 2019) *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power,* Profile Books, London, UK.

Opinions expressed in articles published by AsiaGlobal Online reflect only those of the authors and do not necessarily represent the views of AsiaGlobal Online or the Asia Global Institute

AUTHOR

Sun Sun Lim

Singapore University of Technology and Design (SUTD)

Sun Sun Lim is professor of communication and technology and dean of humanities, arts and social sciences at the Singapore University of Technology and Design (SUTD). She recently authored *Transcendent Parenting – Raising Children in the Digital Age* (Oxford University Press, 2020) and co-edited *The Oxford Handbook of Mobile Communication and Society* (Oxford University Press, 2020). She serves on the boards of the Social Science Research Council, the Singapore Environment Council, the Media Literacy Council, and eleven international journals. She was named to the inaugural Singapore 100 Women in Tech list in 2020 for her pioneering research on the social impact of technology. From 2018-20, she served as a nominated member of the Parliament of Singapore, where she was an advocate for more responsible use of artificial intelligence, transparency in data sharing, greater digital inclusion, and enacting digital rights for children.

Roland Bouffanais

University of Ottawa

Roland Bouffanais is associate professor at the University of Ottawa. His research focuses on the interdisciplinary intersections of complexity, network science, control theory, machine learning, and multi-agent systems. He has published over 100 peer-reviewed papers in top scientific journals and conference proceedings. He authored *Design and Control of Swarm Dynamics* (Springer, 2016). He received his PhD in computational science from École polytechnique fédérale de Lausanne (EPFL) in Switzerland, for which he was awarded the ERCOFTAC (European Research Community