STATUTORY REVIEW OF THE ONLINE SAFETY ACT 2021

ORYGEN SUBMISSION

Orygen welcomes the opportunity to provide a submission to the Statutory Review of the Online Safety Act 2021. The submission addresses the potential for additional arrangements to increase the safety of young people using social media.

This submission was written on the lands of the Wurundjeri people of the Kulin Nation. Orygen acknowledges the Traditional Owners of the lands we are on and pays respect to their Elders past and present. Orygen recognises and respects their cultural heritage, beliefs and relationships to Country, which continue to be important to the First Nations people living today.

ABOUT ORYGEN

Orygen is the world's leading research and knowledge translation organisation focusing on mental ill-health in young people. At Orygen, our leadership and staff work to deliver cutting-edge research, policy development, innovative clinical services, and evidence-based training and education to ensure that there is continuous improvement in the treatments and care provided to young people experiencing mental ill-health.

Orygen conducts clinical research, runs specialist clinical services, including five headspace centres, supports the professional development of the youth mental health workforce and provides policy advice relating to young people's mental health. Our current research strengths include: early psychosis, mood disorders, personality disorders, functional recovery, suicide prevention, online interventions, neurobiology and health economics.

CONTACT DETAILS
For further information, please contact:

David Baker
Manager, Policy

THE MENTAL HEALTH IMPACTS OF RECOMMENDER SYSTEMS

This submission is responding to Term of Reference 4: whether additional arrangements are warranted to address online harms not explicitly captured under the existing statutory schemes, including: other potential online safety harms raised by a range of emerging technologies, including but not limited to: recommender systems. Suicide is the leading cause of death among young Australians under 25, and rates of self-harm and suicide-related thoughts and behaviours among this demographic have been steadily increasing in recent decades.[1] The high frequency and duration of social media usage among young Australians has prompted concerns about the role of social media in introducing or worsening negative mental health outcomes for young people.[2]

A central factor underpinning concerns about the harms of social media is the significant amount of unregulated and uncensored content that young people are exposed to daily, including graphic images of self-harm and suicide. Although some young people actively seek out this content, it often appears on their newsfeeds unsolicited.[3] Young people who are exposed to suicide are more likely to develop suicidal thoughts, feelings, and behaviours and it is thought that exposure to graphic information plays a role in suicide imitation or contagion. Therefore, the volume, nature, and frequency

REVOLUTION IN MIND

of exposure to distressing content on social media is a significant social problem that jeopardises the online safety of young people and warrants concern.

Recommender systems play a central role in the spread and reach of potentially harmful online content. These systems, aimed at maximising engagement with social media platforms can inadvertently lead to promotion of sensitive, graphic, or harmful content. This is particularly concerning for young people who are already distressed and are likely to become over-exposed to harmful content that amplifies or exacerbates their self-harm or suicide ideation. For instance, a study conducted in the United Kingdom reported that in a sample of 5,294 young people, 75 per cent had viewed self-harm content online before the age of 14 years.(3) Of those who encountered self-harm and suicide-related content, 83 per cent reported that they had not intentionally searched for it; many reported that it worsened their mood, and over three quarters reported that they self-harmed in similar, or a more severe way, after exposure to the content. Other reports suggest that on TikTok, teenagers are exposed to mental health content every 39 seconds, with content related to suicide every 2 minutes and 36 seconds.(4) Distressing content is some of the most highly viewed content online,(5) and despite repeated calls for regulation, recommender systems enable this content to spread quickly.

A growing body of evidence recognises that although social media itself is not inherently harmful, users can experience benefits or risk of harm through social media, depending on a range of factors. Indeed, several mental health and suicide prevention interventions have underscored the potential for social media to be utilised as a beneficial tool for social connection and support, and to spread protective information and resources quickly to large audiences.(1, 6-9) However, the benefit and protective utility of social media in these efforts does not negate serious concerns relating to the reach and impact of graphic or harmful content that young people are exposed to regularly on social media.

Recommendation 1

That recommender systems should not promote self-harm and suicide content, especially for users under the age of 25. This recommendation does not suggest a uniform ban on all self-harm or suicide related content, but rather this content should not be automatically recommended to a user. Mechanisms for regulating algorithms/recommender systems of social media platforms is required, with consequences for violation.

In instances where young people are utilising social media to seek support or information about their experiences related to self-harm and suicide, recommender systems present a unique opportunity to promote protective and preventative content to these users. Indeed, anyone who is searching for, or engaging with, content related to self-harm or suicide for any reason is likely to benefit from exposure to safe and protective content. Evidence shows that suicide prevention information delivered via social media is safe, and may present a range of benefits such as increased knowledge and confidence discussing such topics.(1, 6) The potential for recommender systems to enable suicide prevention information to reach young people in a timely way with protective and helpful information is underexplored. More research is needed on real-time and just-in-time supports that could be delivered by trusted sources, alongside considerations of acceptability by young people and issues of privacy.

Recommendation 2

With respect to user privacy, the potential for recommender systems to be used beneficially by identifying users in distress and providing timely support or intervention should be better explored in partnership with social media platforms. If a user is detected as experiencing distress (including risk of self-harm and suicide), recommender systems should be used to promote/deploy evidence-informed content that is protective (for example, #chatsafe).

REFERENCE LIST

1.   La Sala L, Teh Z, Lamblin M, Rajaram G, Rice S, Hill NTM, et al. Can a social media intervention improve online communication about suicide? A feasibility study examining the acceptability and potential impact of the #chatsafe campaign. Public Library Science 2021.

2.   Centre for Digital Wellbeing. The Impacts of Social Media in Australia - Research Brief Kingston, ACT: Centre for Digital Wellbeing; 2021.

3.   Samaritans. How social media users experience self-harm and suicide content. Surrey, United Kingdom; 2022.

4.   Center for Countering Digital Hate. Deadly by design. 2022.

5.   O'Connor C. Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok. Institute for Strategic Dialogue 2021.

6.   La Sala L, Pirkis, J., Cooper, C., Hill, N. T. M., Lamblin, M., Rajaram, G., Rice, S., Teh, Z., Thorn, P., Zahan, R., & Robinson, J,. Acceptability and Potential Impact of the #chatsafe Suicide Postvention Response Among Young People Who Have Been Exposed to Suicide: Pilot Study. 2023.

7.   Robinson J, Hill NTM, Thorn P, Battersby R, Teh Z, Reavley NJ, et al. The #chatsafe project. Developing guidelines to help young people communicate safely about suicide on social media: A Delphi study. PLoS ONE. 2018;13(11):1-13.

8.   Robinson J, Thorn P, McKay S, Hemming L, Battersby-Coulter R, Cooper C, et al. #chatsafe 2.0. updated guidelines to support young people to communicate safely online about self-harm and suicide: A Delphi expert consensus study. PLoS ONE. 2023;18(8):1-17.

9.   Robinson J, Thorn P, McKay S, Richards H, Battersby-Coulter R, Lamblin M, et al. The steps that young people and suicide prevention professionals think the social media industry and policymakers should take to improve online safety. A nested cross-sectional study within a Delphi consensus approach. Frontiers in Child and Adolescent Psychiatry: Frontiers Media S.A.; 2023.