



Suicide Prevention  
Australia



# AI and Digital Safety

JULY 2025



# Suicide Prevention Australia

Suicide Prevention Australia is the national peak body for the suicide prevention sector. We exist to provide a clear, collective voice for suicide prevention, so that together we can save lives.

We represent more than 350 members ranging from national household name agencies to small community-based organisations and local collaboratives in every State and Territory, as well as individual service providers, practitioners, researchers, students and people with lived experience. This includes more than 140,000 employees and volunteers across Australia.

We believe that through collaboration and shared purpose, we can work towards our ambition of a world without suicide.

# Contents

<b>Executive Summary</b>	<b>4</b>
<b>Introduction</b>	<b>5</b>
<b>Key Concepts</b>	<b>6</b>
Digital Safety	6
Artificial Intelligence (AI)	6
<b>Methodology</b>	<b>7</b>
<b>Benefits of Digital Spaces and Tools</b>	<b>9</b>
Community	9
Benefits of Digital Spaces and Tools	9
Diversity and Acceptance	10
Information Accessibility	10
Identifying People at Risk	11
Support for Those Exposed to Suicide	12
Augmenting Support Services	13
<b>Negative Impacts of AI and Digital Media</b>	<b>14</b>
Misinformation, Discrimination and Stigma	14
Bullying and Harassment, Image-Based Abuse	16
Isolation and Loneliness	17
Financial distress	18
Lack of Transparency	19
<b>Conclusion</b>	<b>20</b>
1: Require a safety-by-design approach	20
2: Enforce Transparency	22
3: Co-design of Policies Surrounding Digital and AI Safety	23
<b>References</b>	<b>24</b>



# Executive Summary

Rapid development in the uses of digital platforms, and in artificial intelligence (AI), present both increased risks and potential benefits across a range of areas, including suicide prevention. Combining insights from a roundtable discussion of representatives from key organisations in this space, with research on digital and AI safety, this paper identifies key issues and provides recommendations.

The key themes identified are:

## Benefits of digital spaces and tools

- Community
- Diversity and acceptance
- Information accessibility
- Identifying people at risk
- Support for people exposed to suicide
- Augmenting support services

## Negative impacts of AI and digital media

- Discrimination and stigma
- Bullying and harassment
- Isolation and loneliness
- Financial distress
- Lack of transparency

Analysing insights from the roundtable and available research on these themes, three recommendations were developed:

### 1. Require a safety-by-design approach

Federal and State/Territory governments require social media platforms and AI companies to demonstrate a safety by design approach which puts the onus and regulatory requirements on tech companies to create safe online environment.

### 2. Enforce transparency

Federal government develops a transparency accreditation scheme to support technology companies to ensure transparency with respect to social media algorithms and AI model training.

### 3. Co-design of policies surrounding digital and AI safety:

Governments engage in genuine co-design that ensures a range of diverse perspectives to contribute to all policies that impact on suicide risk related to digital and AI spaces. This includes partnering with Aboriginal and Torres Strait Islander experts to co-design culturally relevant digital content and safety features and involving young people and LGBTIQ+ individuals.



# Introduction

Accessing and utilising digital platforms has become common place in the modern world. With mobile phones with processing speeds around 5,000 times faster than supercomputers from the 1980s<sup>1</sup>, users of all ages can create and engage with content online.

Together with the rise of smartphones, social media has become ubiquitous in everyday life. It is estimated that around 82.5% of Australia's total internet user base used at least one social media platform in January 2024.<sup>2</sup>

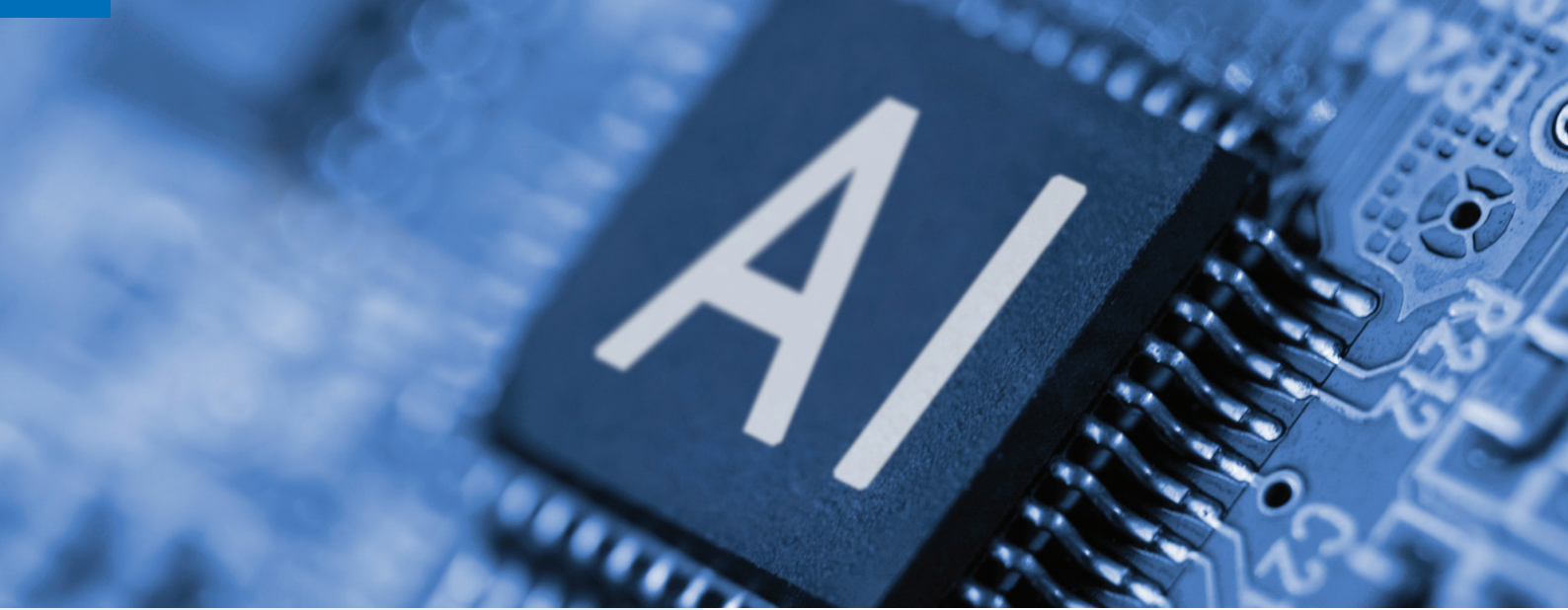
Furthermore, advancements in artificial intelligence (AI) have made it more and more accessible. Australia alone had 38 million searches for ChatGPT and Gemini, both artificial intelligence chatbot tools, indicating Australians' curiosity and appetite for AI technologies.<sup>3</sup>

The rise in social media usage and AI tools has presented the sector with unique challenges that it must be equipped to deal with. In Suicide Prevention Australia's Community Tracker, it was reported that more than a third (36%) of young Australians (aged 18-24) have experienced elevated distress levels due to social media, self-image, and bullying in the past 12 months during the September 2024 quarter. This is the highest it has been since the same quarter in 2022.<sup>4</sup>

In the most recent March 2025 quarter, this figure dropped to around one in four (24%), resulting in an 11-quarter average of 26%. In addition, research undertaken by the University of Queensland reveals that less than half of Australians are comfortable with and trust the use of AI at work, and only a minority of Australians believe the benefits of AI outweigh the risks.<sup>5</sup>

However, digital platforms also provide positive mental health benefits through facilitating online help-seeking where individuals cite an increase in the sense of control over their help-seeking journey, meeting the needs of those with a preference for self-reliance, or as a gateway for further help-seeking.<sup>6</sup>

These developments bring both increased risks and potential benefits and so present a unique opportunity for the suicide prevention sector. The opportunity exists to promote digital and AI safety measures that ensure Australians use these platforms with confidence to navigate potential harms and to help ensure these tools are used effectively to support those most at risk.



# Key Concepts

---

## What is Digital Safety?

Digital safety includes being aware of the nature of the possible threats that one could encounter while engaging in activity through the internet, according to UK-based not-for-profit SWGfl. These could be security threats, protecting and managing your personal data, online reputation management, and avoiding harmful or illegal content.<sup>7</sup>

At a more technical level, digital safety is defined as a branch of cyber security that deals with people and the levels of online comfort, convenience, and privacy. It includes being aware of the risks associated with your online activity and avoiding being exposed to unwanted information, materials, or risks. Common threats to online safety include phishing, scams, malware, cyberstalking, cyberbullying, online predators, and sextortion.<sup>8</sup>

In a suicide prevention context, digital safety can be viewed as practices and measures taken to protect individuals from undue psychological harm or to prohibit the promotion of suicide, or content which glamorises suicide. These practices include the responsible design and use of online tools such as mental health apps and crisis hotlines, among others. This extends not only to the protection of individuals who are at risk of suicide but the mitigation of creating new risks for all users. With the increasing use of AI across a range of digital contexts, it must also be considered in digital safety.

## What is Artificial Intelligence?

According to the Australian Cyber Security Centre, AI is the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making and translation between languages.<sup>9</sup> Today, it's used most often to refer to generative AI models such as ChatGPT. In addition to text, generative AI can also be used to create videos or audio.



# Methodology

Suicide Prevention Australia, with the support of KPMG, invited member organisations operating within digital spaces to a roundtable discussion on their hopes and fears for these technologies. Representatives from batyr, Everymind, Gayaa Dhuwi (Proud Spirit) Australia, LGBTIQ+ Health Australia, Lifeline Australia, ReachOut, and SANE participated in the roundtable which, along with the research referenced, forms the basis of this paper.

Participants were provided a discussion paper that outlined key definitions and recent developments in the digital and AI space, ahead of the roundtable. The roundtable discussions focused on the following key questions:

- What are the risks and benefits of age-based restrictions on the use of social media?
- What other ways can social media platforms (and other platforms with social aspect such as some messaging services and online games) be made safer in regard to suicide risks, particularly for vulnerable populations such as children and young people?
- What are the potential risks and benefits of integrating AI in suicide prevention strategies, particularly concerning vulnerable populations?

- How can AI services and AI-driven platforms better identify and respond to signs of suicidal ideation while respecting user privacy and autonomy?
- What actions should the suicide prevention sector take to mitigate suicide and self-harm content in digital spaces as well as suicide methods online?
- What actions should government be taking to mitigate suicide and self-harm content in digital spaces as well as suicide methods online?

Participants were then invited to list their hopes for digital environments, their concerns, as well as what they thought were the strengths of AI and the digital world. An open discussion followed to develop these ideas and highlight which concepts resonated with members of the roundtable.

Guided by these concepts or ideas, attendees then broke out into groups and were provided with questions to guide collective responses. To flesh out the recommendation, the following questions were also asked:

- In relation to this recommendation, what do we see in our field today?
- What are the underlying challenges that this recommendation is intended to address?



- How can this recommendation facilitate/enable conversations with our various stakeholders?
- What is the spirit/intent that sits behind this recommendation and the meaning we give to week words?
- In the future, how might this come to life for an individual person – what would make this more than words on a page?
- In the future, how might this come to life for our society?
- In relation to this recommendation, what is our vision for our field? What is the meaningful change that we would like to see shaped by this recommendation?
- What are we seeking to move away from?
- What are we seeking to move towards?
- What is the value of this recommendation to individuals, society, our field, and to government?

Following the roundtable, feedback was analysed from the discussion and condensed into themes that guide this paper. The key themes that came out of the discussion were as follows:

### **Benefits of digital spaces and tools**

- Community
- Diversity and acceptance
- Information accessibility
- Identifying people at risk
- Support for people exposed to suicide
- Augmenting support services

### **Negative impacts of AI and digital media**

- Discrimination and stigma
- Bullying and harassment
- Isolation and loneliness
- Financial distress
- Lack of transparency

Online desktop research that focused on these themes, along with current events relating to digital and AI safety and its links to suicide and suicide prevention, was then conducted. Lastly, recommendations were developed based on available information and previous recommendations provided by our members.



# Benefits of Digital Spaces and Tools

---

It cannot be understated how much the internet, social media, and digital spaces and tools have shaped the sector. They lower barriers to entry, and real-time nature of digital spaces were cited as the key benefits for users reaching out for support online. Information can also be easily accessed by anyone with a device and an internet connection. As one roundtable participant stated, there were “endless possibilities”.



## Community

Digital spaces provide users with means of accessing community when it is difficult to find in person and has provided a voice for those who feel that no one listens. Roundtable participants shared that they continue to see people connect and support each other through different social media platforms. One way that elevated the sense of community was storytelling and how posts online can help individuals feel less alone, even if they are physically isolated. Recent evidence points to the value in appropriate storytelling, often called narrative psychology to share stories of hope, and to promote help-seeking for people in suicidal distress.<sup>10</sup>

One example of this is that Beyond Blue hosts online forums that are free, anonymous, and moderated where people can connect to share their experiences with mental health challenges.<sup>11</sup> It offers a safe, supportive space for users to seek advice, give support, and feel less alone in their struggles.

Whilst international research shows that participating in such forums can significantly reduce feelings of isolation, encourage people to access additional mental health services, and inspire positive lifestyle changes.<sup>12</sup> By fostering peer-to-peer connection and offering a judgment-free environment, the forums play a vital role in promoting mental wellbeing, especially for those who may be hesitant to seek face-to-face help.



## Diversity and Acceptance

In a position paper released by Gayaa Dhuwi, the organisation highlighted the positive effects of social media on Aboriginal and Torres Strait Islander youth. It provides important spaces for them to express and celebrate their cultural identity, helping to strengthen pride and connection to heritage. It also enables young people to stay connected with family and community members across long distances, maintaining vital support networks. Additionally, by engaging with digital platforms, children and young people can build valuable digital skills that may support their future education and employment opportunities.<sup>13</sup>

Research funded by Australian Rotary Health suggests that higher levels of cultural engagement such as speaking Indigenous languages, cultural connectedness and participating in cultural events, reduces risk factors for youth suicide. Gibson and colleagues found that the youth suicide rate was 44% lower in areas with greater community participation in cultural events, described by the authors as “cultural social capital”.<sup>14</sup>



## Information Accessibility

Mental health NGO ReachOut’s ‘Ask ReachOut’ uses AI to provide accessible, personalised, reliable information and support by matching users’ questions with answers from a curated database of ReachOut’s articles and resources. The issues and questions answered by Ask ReachOut include common struggles young people face such as building better study habits, managing anxiety, and getting free mental health support. These responses are carefully researched by ReachOut staff and shaped by user feedback, ensuring they are relevant and trustworthy. ReachOut is committed to improving Ask ReachOut by expanding its library and covering more topics to better meet users’ needs.<sup>15</sup>

Accessible mental health resources can contribute to higher levels of mental health literacy which has proven to help with one’s mental health status, acquiring knowledge of the available services, improving help-seeking behaviours and attitudes, and having the mental health first aid skill confidence in supporting peers with mental health problems.<sup>16</sup>

**“We have all these other incredible fields using data, using artificial intelligence, using community – so if we want to solve the public health issue that suicide is, we need to do it together”**

- Dr Sandersan.

## Identifying People at Risk

The roundtable participants suggested that utilising AI in behavioural pattern recognition and insight generation around self-harm and suicidal behaviours, is one beneficial use of AI. There is evidence suggesting that AI can be used to detect subtle data patterns, relationships, and variable interactions to determine suicide risk. Data can be pulled from different sources and synthesised as well as scaled to continuously improve.<sup>17</sup>

The Black Dog Institute has done work on investigating the evidence for machine-learning models and their ability to predict future suicidal behaviours and thoughts as compared to traditional models. These models typically include risk assessment tools used in emergency departments such as questionnaires and rating scales. Unfortunately, evidence suggests existing traditional suicide risk assessment tools are ineffective in accurately predicting suicide risk in practise. Their evaluation of 54 machine-learning algorithms found that they outperformed traditional risk prediction models in predicting suicide-related outcomes, but more research is necessary to improve and validate these algorithms.<sup>18</sup>

Another way data could be used is to analyse search terms. Dr Onie Sandersan, research fellow at the Black Dog Institute in Sydney, observed how a banner with the hotline for the national suicide prevention line often would come up on online search results, particularly if someone was actively looking for suicide supports. He also found that some keywords, such as those referring to loneliness or hopelessness, did not trigger a link to the hotline and that some generated links were potentially harmful. For example, when he typed in keywords relating to a well-known Australian location where suicides occur, “instead of giving me the hotline number, it gave me a map,” he told the British Medical Journal.


Dr Sandersan also found the risk was not confined to Australia, with a search on Indonesian Google generating tips on ‘how to end your life’ rather than links to support.<sup>19</sup>

There are risks and limitations to the work, of course, including that privacy regulations mean the researchers cannot know whether people who click onto the landing pages go on to access support and avoid suicide.

However, Dr Sandersan, whose original research was funded by the Suicide Prevention Australia Research Fund, believes a key lesson is that answers for suicide prevention will not come from just within the suicide prevention sector.

The research continues to grow, and plans are underway to use this approach across multiple platforms, so that every person who is in distress and engages in online behaviours indicative of distress will be met with the help that they want and need. To that end, a summit co-led by Dr Sandersan sought to bring the industry together to understand how the suicide prevention sector can utilise multiple platforms to create a safety net or ecosystem that meets people where they are.

Dr Sandersan said he was amazed that the simple question ‘can we run some ads on Google to see if we can help people who are suicidal?’ has managed to “bring the world’s major tech companies together, for the first time, to see whether we can have a healthier online ecosystem that is safe for everyone regardless of where they are in their point of distress”.<sup>20</sup>



**“The research provided evidence that #chatsafe was a safe and potentially efficacious intervention that could form part of future postvention responses and help reduce the risk of imitation or contagion after a suicide has occurred,”**

*- Professor Robinson.*



## Support for Those Exposed to Suicide

Another benefit of digital tools is providing different types of support services in real-time. An example highlighting this is a research project involving 260 people aged 16 to 26 years who had been recently exposed to suicide. In an international first, Professor Jo Robinson, who leads the suicide prevention research unit at youth-focused mental health organisation, Orygen, led a trial testing the #chatsafe intervention, through platforms like Instagram, Facebook and Snapchat, to support young people to communicate safely online about self-harm and suicide.

Each week over the six-week trial, the researchers sent direct messages of tailored content that had been co-designed with young people to the participants. The content covered safe communication about death and suicide, how to seek help and tips on supporting friends if they were worried about them. Research has shown that many people fear that discussing suicide with a friend or family member, or asking someone if they are experiencing suicidal thoughts, can be harmful, and thus did not know how to safely support a friend or family member who might be at risk of suicide.<sup>21</sup>

Many participants reported increased willingness to intervene because of receiving the #chatsafe content and said their perceived internet self-efficacy, confidence and safety when communicating about suicide on social media had improved. Importantly, there was no evidence that an online program to support those who experienced grief due to suicide, called postvention, increased the risk of suicide.

Backed by their findings, the research team has been working with health departments and Primary Health Networks to conduct more than 20 postvention campaigns to date across Australia and New Zealand, aimed at specific communities where a suicide has occurred.

“We are notified when a young person dies by suicide, so we can roll out a #chatsafe campaign to the young people in the community impacted by that death, as well as adult-facing resources for parents and carers, and specific resources for educators,” Professor Robinson said. So far, the campaigns have registered 46 million ‘impressions’ (the number of times their material shows up on a news feed), reaching more than 6.5 million individuals and generating 250,000 ‘click throughs’ to the #chatsafe website for more information.<sup>22</sup>



## Augmenting Support Services

More innovative solutions were also considered at the roundtable, with advancements in technology allowing for new and cutting-edge solutions such as individuals looking to AI chatbots as therapists. Participants cited low barriers to entry due to cost and timeliness as the main positives.

In research conducted by Orygen, out of 107 community members, around a third used AI tools, most commonly ChatGPT, for quick emotional support or in some cases, as their own personal 'therapist'. Further, 76.7% of community members reported AI to be beneficial to varying degrees.<sup>23</sup>

This is supported by the views of Professor Jill Newby of the Black Dog Institute. In her interview with the University of New South Wales, she said advice given by AI chatbots is based on cognitive behavioural therapy which may be beneficial to those who find the in-person version effective. However, she says that a healthy level of scepticism is necessary when engaging with these chatbots. In ChatGPT's own words, "ChatGPT may miss subtle emotional cues or overgeneralise advice. It cannot replace professional mental health evaluation or diagnosis."<sup>24</sup>

However, nearly half of the community members from the Orygen study also experienced risks or harms when using AI such as issues regarding data privacy and unhelpful advice, among others.<sup>25</sup> This type of use of AI chatbots highlights the need for these platforms to provide accurate information regarding which helplines or services one could access when experiencing distress.

Support could also be enhanced for specific communities such as those of culturally and linguistically diverse backgrounds. Thoughtworks, a global tech consultancy, did a proof-of-concept in 2021 using technology that automates translation in clinical healthcare settings in a culturally sensitive manner that enhances patient experience.<sup>26</sup>



## Negative Impacts of AI and Digital Media

Roundtable participants identified several negative issues and trends which have emerged or are exacerbated through the widespread use of AI and digital media. These included fake news and misinformation, hate speech, cyberbullying, harassment, sexual harassment, and concerns around privacy and the misuse of data.

These negative impacts of social media and AI can often interact with and exacerbate the socio-economic and environmental determinants of suicide. These are the social, economic and environmental conditions which can increase a person's risk of suicide. Suicide Prevention Australia's background paper on this discussed several determinants which are particularly relevant to AI and digital safety, including: discrimination and stigma, bullying and harassment, financial distress, isolation, and loneliness.<sup>27</sup>



## Misinformation, Discrimination and Stigma

Participants expressed their concerns with the discrimination and stigma that can arise because of misinformation and fake news. In February 2025, media watchdog NewsGuard identified 1,254 unreliable AI-generated news and information websites spanning 16 languages compared to around 200 sites at the start of 2023.<sup>28</sup> Meta CEO Mark Zuckerberg has also said that his platforms (Facebook, Instagram, Threads) will be abandoning their third party fact-checking program starting in the United States of America, and then drop fact checking in other countries after 2025.<sup>29</sup>

There is strong evidence suggesting that all forms of stigma, racism, discrimination, and prejudice have links to suicide risk. Among lesbian, gay and bisexual youth, stigma and discrimination experiences are associated with increased suicidality internationally.<sup>30</sup>

In Europe, disinformation and hate speech by some politicians, have led to violence against LGBTQIA+ people reaching its highest point in the past decade.<sup>31</sup> Thus exposure to hate speech and misinformation through social media, can heighten the risk of suicide, for this cohort, as it can carry the real threat of physical violence, which is already has a higher risk according to AIHW.<sup>32</sup>

LGBTIQ+ Health Australia (LHA) is acutely aware of the negative effects of misinformation, as well as hate speech, which can cause delay in seeking general healthcare and gender-affirming care for LGBTQIA+ young people.<sup>33</sup>



**“The more common locations were online and media particularly social media, followed by public places... Workplaces - in this report it’s still the highest area of self-reports by First Nations people.”**

*- Report Author, Professor Allison.*

They argue that timely access to evidence-based gender-affirming care reduces suicide risk, but a lot of misinformation in the public arena is fuelling stigma, fear, and confusion. LHA highlights that this misinformation often misrepresents gender-affirming care, discouraging young people from seeking help and increasing mental health risks. They stress the need for accurate information and respectful public discourse to support timely, life-saving care.

Not only are some young people being denied healthcare, but some are also choosing to not seek professional care due to misinformation. A global survey by Edelman in 2025 revealed that Gen Z adults under 35 are increasingly turning to peers rather than medical professionals for health advice. This trend, intensified by the COVID-19 pandemic, has led nearly 60% of young adults to regret at least one health decision based on inaccurate or misleading information.

Other vulnerable cohorts who might experience hate speech, discrimination and racism due to social media and AI, include first nations people.<sup>34</sup> One study found that younger Aboriginal people aged 16-20 experiencing racism had increased suicide risk.<sup>35</sup>

There were numerous reports of increased racism and violence following the Referendum on the Voice to Parliament, according to the report by the Jumbunna Institute.<sup>36</sup>

The report found that 35% of respondents highlighted aggressively racist behaviour: physical violence, verbal abuse, hate speech, threats, intimidation, bullying, and property damage. Report author, Professor Allison says it was hard for First Nations people to avoid being abused or stereotyped, in the wake of the Referendum.

According to Dr Terri Janke, an international authority on Indigenous Cultural and Intellectual property, racist stereotypes and attitudes could be perpetuated by AI due to AI getting information from systems that have inherent and structural racism.

At the roundtable, there was strong support for the necessity of using First Nation’s strategies on wellbeing and suicide, as well as fostering collaboration between technology developers and Aboriginal and Torres Strait Islander leaders to create safer digital environments that prioritise the mental Health of all Australians.

Exposure to both positive content and harmful content around mental health and suicide was another issue raised in the Roundtable discussions. Approximately 1-in-4 teenagers are exposed to content around suicide before the age of 14, according to recent research.<sup>37</sup> The same study found that young people are exposed to mental health information on average every 39 seconds while online. One of the lead researchers in the study, Dr Louise Da Sala points out that for young people social media platforms can both mitigate and amplify the risk of suicide. For instance, while social media can increase belonging, reduce a sense of being a burden, thus reducing suicidal risk, it can alternatively reduce young people’s sense of belonging, increase their sense of being a burden and promote ideas about the means of suicide.



## **Bullying and Harassment, Image-Based Abuse**

Cyberbullying, or bullying using technology, can lead to victims feeling guilty, hopeless, and depressed. Suicide Prevention Australia's Socioeconomic and Environmental Determinants of Suicide paper summarised the research on the strong links between bullying, including cyberbullying, and risk of suicide and suicidality.<sup>38</sup> A systematic review also found that any participation in bullying increases the risk of suicidal ideation and suicidal behaviour in a broad spectrum of youth.<sup>39</sup>

Another meta-analysis found that any involvement in bullying is associated with suicidal ideation and behaviour.<sup>40</sup> People who experienced cyberbullying-bullying experienced a 2-4 times higher risk of concurrent suicidal ideation or attempt compared to those who had not been victimised online, even after adjusting for face-to-face victimisation and other key confounders.<sup>41</sup>

In addition, cyberbullying can be more difficult to deal with due to the ease of sharing bullying materials or comments with others, and due to the anonymity of cyberbullying.<sup>42</sup> In 2019, just over 1-in-5 young people aged 15-19 reporting experiencing bullying in the previous 12 months.<sup>43</sup> Individuals from culturally and linguistically diverse backgrounds are also adversely impacted by cyberbullying. In a report released by the NSW Government, children and young people from a refugee or migrant background cited bullying as the top issue that they face.<sup>44</sup>

Cyberbullying can also include using social media or AI for sexual harassment which can lead to suicidal thoughts and behaviours. When AI-generated or manipulated sexual content is shared without consent on platforms like Instagram or TikTok, it can trigger severe psychological trauma, including anxiety, depression, and social withdrawal.<sup>45</sup> The viral nature of social media amplifies humiliation and loss of control, often making victims feel powerless and exposed on a global scale.

In London, a fourteen-year-old girl died by suicide as a result of constant cyberbullying. Her face had been photoshopped onto bodies of pornography performers and had received multiple negative comments on TikTok video she would post.<sup>46</sup>

Generative AI has also been shown to enable sexual harassment through the use of deepfakes, defined by the eSafety Commission as images or videos that have been edited to create extremely realistic but false depictions of people.<sup>47</sup> A 35-year-old schoolteacher had her photos used by a close friend to create sexually explicit content without her consent.<sup>48</sup> In a separate incident, around 50 high school students reported having their photos altered in sexually explicit ways leading to the arrest of a teenage boy.<sup>49</sup>

Another area of concern is the way social media algorithms feed images of unrealistic body ideals to girls and young women. Organisations such as the Center for Countering Digital Hate (CCDH) and investigative journalists have found that girls searching for exercise tips for health, or healthy food are quickly fed links to social media influencers promoting unrealistic body images.<sup>50</sup>

The CCDH cites many cases where girls might be linked to social media accounts promoting anorexia. Anorexia is one of the most dangerous mental health conditions, with the highest rate of mortality for any mental health condition.<sup>51</sup> A 2021 study showed that suicide risk is elevated for women with eating disorders.<sup>52</sup>

Rates of suicidality and self-harm among teenage girls, and young women are extremely high, and have been growing at an alarming rate over the past decade, according to a 2019 study.<sup>53</sup> There is little research on the exact causes, some research points to harmful body image<sup>54</sup>, the high rate of bullying<sup>55</sup> and sexual harassment and abuse as causing this alarming increase in self-harm.<sup>56</sup>



## Isolation and Loneliness

In Suicide Prevention Australia's Community Tracker, December 2024 results revealed that 20% of Australians are experiencing distress at elevated levels compared to the same time last year due to social isolation and loneliness. Multiple studies have shown the links between social isolation and suicide risk, which is also outlined in Suicide Prevention Australia's paper Socio-economic and environmental determinants of suicide: background paper.

One systematic review cited in the background paper found that social isolation and a lack of close friends or family network increases risk of suicide among adolescents.<sup>57</sup> For men, the risk of suicide decreases with increasing social integration. Research has also found that among women high levels of social integration is associated with a lower risk of suicide.<sup>58</sup>

While a retrospective study of coronial data in rural NSW, SA, Qld and Tas from 2010-2015 found that the suicide rate was 11% higher in rural areas and increased with remoteness.<sup>59</sup>

While social media and AI can reduce loneliness and isolation, often creating online communities which are particularly beneficial for marginal groups, and people living in remote areas, it can also create new risks. One of these risks is AI companions, according to parental media monitoring group Common Sense Media and mental health professionals from Stanford Brainstorm Lab.

AI companions are "a new category of technology that goes beyond simple chatbots," according to the Parents' Ultimate Guide to AI Companions and Relationships.<sup>60</sup> They are specifically designed to, among other things, "simulate emotional bonds and close relationships with users, remember personal details from past conversations, role-play as mentors and friends, mimic human emotion and empathy, and "agree more readily with the user than typical AI chatbots," according to the guide.<sup>61</sup> However, AI cannot address the deep human need for reciprocal relationships. It can, unfortunately, create a false sense of closeness, exacerbating isolation.<sup>62</sup> In one recent case this led to the suicide of a teen, who developed an intense relationship with his AI Companion.<sup>63</sup>

**“The human impacts of Robodebt were being reported: families struggling to make ends meet receiving a debt notice at Christmas, young people being driven to despair by demands for payment, and, horribly, an account of a young man’s suicide.”**

- Royal Commissioner, Catherine Holmes.



## Financial distress

The prevalence of online scams and the effects of Robodebt were two major negative impacts of digital and AI tools, identified by the Roundtable participants. Roundtable participants identified how these two issues are linked to financial distress, which is one of the socio-economic determinants of suicide, outlined in Suicide Prevention Australia’s background paper.<sup>64</sup> AIHW data shows that between 2012 and 2016 the likelihood of dying by suicide was higher among Australians in the lowest income group compared to those in the highest income group, showing the link between financial distress and suicide risk.<sup>65</sup>

In 2024 alone, close to 250,000 scams were reported to the National Anti-Scam Centre resulting in losses to close to \$320 million.<sup>66</sup> A type of scam that came up in discussion was sextortion cases where individuals, particularly young men, are coerced to provide sexually explicit material of themselves which scammers use to blackmail them for different sums of money. This was unfortunately the case with a teenager in Victoria who died by suicide following a sextortion scam.<sup>67</sup> A similar plot led to the death of a teenager in NSW who was hounded to death over \$500.<sup>68</sup> Scammers are now also using AI to copy voices of loved ones or prominent celebrities or politicians to trick individuals into sending money. In 2022, almost 240,000 Australia reporting being victims of this type of scam, resulting to financial losses of \$586 million.<sup>69</sup>

Robodebt was also another recent case where automated data-matching was used between income tax and social welfare data where debt notices were being automatically issued to welfare recipients.

This ignored the realities of insecure or casualised employment with more variable income and not consistent with the social security legislative framework which required entitlements to be calculated based on actual fortnightly income.<sup>70</sup>

The Robodebt Royal Commission found at least three known suicides directly linked to Robodebt. The Commission expressed confidence that these were not the only tragedies of this kind.<sup>71</sup>

The Royal Commission also pointed out that although media reports emerged of the terrible impacts of the Robodebt scheme, nobody in government halted it.

“The beginning of 2017 was the point at which Robodebt’s unfairness, probable illegality and cruelty became apparent. It should then have been abandoned or revised drastically, and an enormous amount of hardship and misery (as well as the expense the government was so anxious to minimise) would have been averted. Instead, the path taken was to double down, to go on the attack in the media against those who complained and to maintain the falsehood that in fact the system had not changed at all.” Royal Commissioner Catherine Holmes writes in the final report.<sup>72</sup>

In addition to the very real cases of suicide, this stigmatisation could have led to other forms of trauma, anxiety, distress, and mental ill-health according to the Royal Australian and New Zealand College of Psychiatrists (RANZCP).<sup>73</sup>



## Lack of Transparency

Underpinning the concerns of the sector and highlighted by the examples above, particularly the Robo-debt case, is the lack of transparency around how these AI tools and social media algorithms work. This can significantly erode user trust and erode trust in government services, as in the Robo-debt case. Participants commented on how it may deter individuals from seeking help, reporting concerning content, or engaging with digital interventions during times of crisis.

When these systems clearly explain how they make decisions – such as what data they use and why they offer certain recommendations – users are more likely to feel informed, respected, and in control.

This sense of clarity can foster trust, especially for individuals seeking help during vulnerable moments. As highlighted in *Psychology Today*, users respond more positively to AI systems that provide insight into their inner workings, yet transparency alone is not always enough. Many people remain hesitant to rely on AI for emotionally complex decisions, preferring human judgment for its perceived empathy and ethical reasoning. Therefore, while transparency enhances the credibility and acceptability of mental health AI tools, it must be coupled with thoughtful design, human oversight, and ethical safeguards to truly support users in crisis.<sup>74</sup>



# Conclusion

Based on insights gathered through the roundtable and supporting research, three recommendations have been developed:

## 1. Require a safety-by-design approach

Federal and State/Territory governments could require social media platforms and AI companies to demonstrate a safety by design approach.

The key themes highlight the need for social media platforms and AI companies to comply with an approach that meets Australians' expectations for a safe online environment. Safety by design, also called a duty of care, puts onus on the technology companies to minimise online threats by anticipating, detecting and eliminating online harms before they occur.<sup>75</sup> Technology companies should be required to ensure that the functionality of the features of a digital service is psychologically safe, according to Reset Tech.

The current legislation in Australia doesn't require a safety by design model, but several organisations, and researchers concerned with children's safety and privacy, have called for the government to introduce legislation requiring tech platforms to comply with safety by design regulations, similar to those in the UK and the European Union.

Implementing safety-by-design principles in digital and AI platforms ensures that these technologies prioritize user safety, minimize harm in the way the platforms are designed. Safety by design measures would minimise harms on social media by proactively integrating

protective measures against scams, harassment, and misinformation. One suicide prevention researcher, Associate Professor Zac Seidler, has suggested that Safety by Design would include removing the addictive components.<sup>76</sup>

**“In other words, social media should be designed from the outset to prevent harming users. It may mean the end of addictive features such as infinite scrolling, frequent push notifications, and auto-play videos.”**

*- Associate Professor Zac Seidler*

“Regulators also need the tools and power to hold platforms to account. That includes financial penalties, more transparent reporting from big tech companies, and taking proactive steps to keep harmful material off these platforms – not just taking down content after the fact.”

In November 2024, the federal government announced its plans to impose a 'digital duty of care' on tech companies.<sup>77</sup> The continued development and implementation of a Digital Duty of Care regulatory model to place broad obligations of digital platforms will enable users to navigate digital spaces and tools safely by requiring tech companies to regularly conduct risk assessments to proactively identify harmful content. This includes making the corresponding design changes to recommendation algorithms and the user interface to mitigate potential harms.<sup>78</sup>



In the European Union, where a digital duty of care model is followed, users can submit online complaints directly to tech companies. A Digital Services Coordinator can then investigate further if a tech company refuses to remove content.<sup>79</sup>

Similarly, in 2025, Suicide Prevention Australia put a submission for the Draft Online Safety (Age Restricted Social Media Platforms) Rules 2025. The submission touches on risks prevalent online such as cyberbullying, which has links to suicide.

We called for the following, which were also based on recommendations from the E-Safety Commission, and include:

- Establishing protocols for engaging with law enforcement, support services, and illegal content hotlines.
  - Implementing processes to detect, flag, and remove harmful or illegal content proactively.
  - Conducting risk management and impact assessments to identify and address potential
  - Implement social contracts during user registration outlining safety responsibilities for all parties.
  - Balance security, privacy, and user safety when managing personal data and information.
  - The exemption of services that have the sole or primary purpose of supporting the health of end users should be maintained in the final Bill.
- In determining which services have the primary function to support health and education, Government should engage with the suicide prevention sector to ensure that all digital media and apps that help prevent suicide are included in this category.
  - An additional exemption category be created for social media platforms that can demonstrate a safety by design approach that meets the expectations of Australians on a safe online environment.

This will not only allow young people to access the benefits of social connections through social media in a safe environment but would also provide an incentive for social media platforms to be safer for all users, benefiting all Australians who use social media.<sup>80</sup>

Reset Tech Australia also released a report and policy briefing outlining the necessary elements for effective digital platform regulation.<sup>81</sup> They suggest effective regulation of digital platforms would include:

- Risk assessments. These could be focused on addressing key priority areas, including systems and content. As one example, we would explicitly recommend an assessment around children's best interests to ensure their rights are advanced and to help create harmonisation with the Privacy Act.
- Risk mitigation measures. Platforms must be required to implement reasonable steps to mitigate each risk identified.
- Framework for public transparency – which would include annual public risk assessments.
- Annual independent audits of risk assessments and transparency reports e.g. repositories that document all paid advertising and details in searchable ways.
- Regulator-direct transparency measures.



## 2. Enforce Transparency

In line with a safety-by-design approach is enforcing transparency with respect to social media algorithms and AI model training. This could be achieved through accreditation of technology companies.

According to the Open Compliance and Ethics Group,<sup>82</sup> transparency in the context of AI is an umbrella term that covers multiple aspects:

- **Explainability of AI Models** – the ability of an AI model to be clearly explained in how it works.
- **Data Transparency** – the feature of an AI model having easily accessible information regarding data origins, lineage, quality, and privacy practices.
- **Model Governance and Documentation** – the maintenance of detail records of model creation.
- **Risk and Impact Disclosure** – the transparency of AI models regarding potential risks of use.
- **Bias and Fairness Assessments** – the mitigation of discriminatory impacts.
- **Governance Framework Transparency** – the transparency with respect to organisational decision-making.
- **Stakeholder Communication** – the frequency and openness of communication between stakeholders.

With respect to social media algorithms, transparency refers to a measure of openness about the purpose, structure and underlying actions of the algorithms used to search for, process and deliver information.<sup>83</sup>

When AI-driven tools operate transparently, individuals are more likely to trust and engage with them.<sup>84</sup> Transparency ensures that users understand how AI detects distress signals, what actions it takes when identifying at-risk individuals, and how their data is used, stored, and protected. This reduces fears of privacy violations and builds confidence in the system.

If we intend to use AI technologies to support suicide prevention efforts, the sector must take charge in understanding and developing the regulations required to ensure transparency from social media platforms and AI systems.

Harvard Business School scholar Professor Shoshana Zuboff argues social media platforms such as Instagram, and AI tech giants are using arguments around freedom of speech, access to services, and democracy to resist regulation, while in fact undermining democracy and violating privacy.<sup>85</sup> However perhaps the sector should present arguments around the potential of transparency to reduce the reputational risk and harm to ordinary users. According to the Harvard Business Review, transparency can decrease the risk of error and misuse, distribute responsibility leading to accountability, enable internal and external oversight, and express respect for people and their autonomy.<sup>86</sup>

Reset Tech Australia provides guidance on how Australia can regulate for transparency. Firstly, transparency must be considered as an integral pillar of effective digital regulation which included several regulatory reforms. Secondly, regulatory mandates for publishing transparency reports must include a detailed sets of metrics that provide clarity about definitions, the operations and effectiveness of content moderations systems. Lastly, these transparency reports need to be part of a broader transparency network that includes additional measures like regulator-facing risk assessments that become public, independent audits.<sup>87</sup>

The development of a transparency accreditation scheme could support technology companies to ensure transparency with respect to social media algorithms and AI model training.



### 3. Co-design of Policies Surrounding Digital and AI Safety

Roundtable participants agreed on the necessity of co-designed policies that take input from a range of diverse perspectives that recognise intersectionality and contribute to the design of all policies that impact suicide risk related to digital and AI safety.

Too often governments take tokenistic approaches to co-design where stakeholders are being invited too late into the conversation leading to a top-down decision approach that do not consider different views. The government's approach to banning social media was an instance when many young peoples voices were not being heard, with criticism legislation was rushed through parliament. People from minority groups, such as Aboriginal and Torres Strait Islander people or culturally and linguistically diverse backgrounds also complain about the lack of co-design, in developing services for them.

Any co-design processes should follow well-developed and respected co-design guidelines. Orygen provides guidance on how to do co-design with young people. Core features of effective co-design include valuing lived experience equally with professional expertise, sharing decision-making power, employing a design-led process, and utilising design methods to support active participation.<sup>88</sup>

Implementing co-design effectively involves several key strategies, including organisations:

- **Fostering a safe and supportive environment** where young people feel empowered to contribute genuinely, knowing their input will influence outcomes.
- **Co-designing opportunities** that are integrated throughout all phases of service development, from initial planning to evaluation, allowing for continuous and meaningful engagement.
- **Providing appropriate training and support** to help young participants build the skills and confidence necessary to engage fully in the process.
- **Preparing to act on the contributions of young people**, demonstrating that their involvement leads to tangible changes, thereby reinforcing the value of their participation.

A co-design process should and could be used to develop AI tools which were designed to proactively detect distress signals in text-based interactions, search queries, or social media activity and trigger appropriate interventions. This approach shifts responsibility to service providers, ensuring that crisis support, such as access to trained professionals or emergency helplines, is readily available rather than leaving users to navigate distressing situations alone, or to navigate to service providers.

# References

- 1 Team, A. A. (2022, November 8). Fast-forward — comparing a 1980s supercomputer to a modern smartphone | Adobe Blog. <https://blog.adobe.com/en/publish/2022/11/08/fast-forward-comparing-1980s-supercomputer-to-modern-smartphone>
- 2 Kemp, S. (2024, February 20). Digital 2024: Australia — DataReportal — Global Digital Insights. DataReportal — Global Digital Insights. <https://datareportal.com/reports/digital-2024-australia#:~:text=More%20broadly%2C%2082.5%20percent%20of,while%2049.3%20percent%20were%20male>
- 3 EducationDaily, & EducationDaily. (2024, October 15). Australia is a world leader in AI-addiction. EducationDaily. <https://educationdaily.au/technology/is-australia-too-ai-addicted/>
- 4 Suicide Prevention Australia Community Tracker. (2024, September). <https://www.suicidepreventionaust.org/community-tracker>
- 5 Trust in Artificial Intelligence: A global study. (2021, July 22). Ai.uq.edu.au. <https://ai.uq.edu.au/project/trust-artificial-intelligence-global-study>
- 6 Pretorius, C., Chambers, D., & Coyle, D. (2019). Young people, Online Help-Seeking and Mental Health Difficulties: A Systematic Narrative Review (Preprint). *Journal of Medical Internet Research*, 21(11). <https://doi.org/10.2196/13873>
- 7 Designed and Developed by Will Earp, South West Grid for Learning Trust Ltd. (n.d.). What is Online Safety? | SWGfL. <https://swgfl.org.uk/online-safety/what-is-online-safety/>
- 8 Sadiku, M. N. O., Ashaolu, T. J., Ajayi-Majebi, A., & Musa, S. M. (2021). Digital safety. *International Journal of Scientific Advances*, 2(5). <https://doi.org/10.51542/ijscia.v2i5.21>
- 9 An introduction to artificial intelligence | Cyber.gov.au. (n.d.). <https://www.cyber.gov.au/resources-business-and-government/governance-and-user-education/artificial-intelligence/an-introduction-to-artificial-intelligence>
- 10 Why storytelling matters in Suicide Prevention | Mental Health Academy. Accessed at: <https://www.mentalhealthacademy.com.au/catalogue/courses/the-eye-of-the-survivor-why-storytelling-matters-in-suicide-prevention>
- 11 See: <https://forums.beyondblue.org.au/t5/custom/page/page-id/Forums-Guidelines>
- 12 Kilpatrick S, Emery S, Farmer J, Kamstra P. Wellness Impacts of Social Capital Built in Online Peer Support Forums. *Int J Environ Res Public Health*. 2022 Nov 22;19(23):15427. doi: 10.3390/ijerph192315427. PMID: 36497502; PMCID: PMC9738771. Accessed at: Wellness Impacts of Social Capital Built in Online Peer Support Forums - PMC
- 13 Gayaa Dhuwi (Proud Spirit) Australia. (2022). Impact of social media on children's development and social and emotional wellbeing. In Gayaa Dhuwi (Proud Spirit) Australia (pp. 1–9) [Report]. <https://www.gayaadhuwi.org.au/wp-content/uploads/2024/11/Impact-of-Social-Media-on-Children-Position-Paper-Board-Endorsed.pdf>
- 14 Gibson M, Stuart J, Leske S, et al. Suicide rates for young Aboriginal and Torres Strait Islander people: the influence of community level cultural connectedness. *Med J Aust* 2021; 214: 514– 518.
- 15 ReachOut. Ask ReachOut | ReachOut Australia. <https://au.reachout.com/ask-ro>
- 16 Bennett, H., Allitt, B., & Hanna, F. (2023). A perspective on mental health literacy and mental health issues among Australian youth: Cultural, social, and environmental evidence! *Frontiers in Public Health*, 11. <https://doi.org/10.3389/fpubh.2023.1065784>
- 17 Araujo, T., Ceron, W., & van der Meer, T. G. L. A. (2024). The moral gatekeepers of AI: How social media users deal with the ethical implications of algorithmic decision-making. *Journal of Business Research*, 173, 114535. <https://doi.org/10.1016/j.jbusres.2024.114535>
- 18 Artificial intelligence may improve suicide prevention in the future - Black Dog Institute | Better Mental Health. (2022, October 21). Black Dog Institute | Better Mental Health. <https://www.blackdoginstitute.org.au/news/artificial-intelligence-may-improve-suicide-prevention-in-the-future/>
- 19 Suicide Prevention Australia. (2024). Insight to Impact: How the Suicide Prevention Research Fund creates knowledge to save lives
- 20 Suicide Prevention Australia. (2024). Insight to Impact: How the Suicide Prevention Research Fund creates knowledge to save lives
- 21 Kenny, B., La Sala, L., Gao, C., Cooper, C., McKay, S., Sabo, A., & Robinson, J. (2025). Young People's Behavior, Confidence, and Safety when Communicating Online about Self-Harm and Suicide: Development and Validation of the #Chatsafe Online Safety Scale. *Cyberpsychology, Behavior, and Social Networking*. <https://doi.org/10.1089/cyber.2024.0558>
- 22 Büyüktür, A. G., & Serçe, F. Ç. (2023). A review of digital mental health services in suicide prevention: Technological and human-centered design perspectives. *JMIR Human Factors*, 10(1), e44535. <https://doi.org/10.2196/44535>
- 23 New study reveals Australians turning to AI for mental health support. (n.d.). Orygen. <https://www.orygen.org.au/About/News-And-Events/2024/New-study-reveals-Australians-turning-to-AI-for-me>
- 24 Newby, P. J. (n.d.). Could you replace your therapist with an AI chatbot? UNSW Sites. <https://www.unsw.edu.au/newsroom/news/2025/03/therapist-as-AI-chatbot>
- 25 New study reveals Australians turning to AI for mental health support. (n.d.-b). Orygen. <https://www.orygen.org.au/About/News-And-Events/2024/New-study-reveals-Australians-turning-to-AI-for-me>
- 26 How AI could bridge the linguistic inclusion gap. (n.d.). Thoughtworks. <https://www.thoughtworks.com/en-au/insights/blog/machine-learning-and-ai/how-ai-could-bridge-the-linguistic-inclusion-gap>
- 27 Suicide Prevention Australia. (2023) The Socio-economic and environmental determinants of suicide. Accessed at: SPA-SEDS-Background-Paper-August-2023-Designed.pdf
- 28 NewsGuard. (2025, February 10). Tracking AI-enabled misinformation: over 1200 'Unreliable AI-Generated news' websites (and counting), plus the top false narratives generated by artificial intelligence tools - NewsGuard. <https://www.newsguardtech.com/special-reports/ai-tracking-center/>

- 29 Riedlinger, M., Watt, N., & Montaña-Niño, S. (n.d.). Meta is abandoning fact checking – this doesn't bode well for the fight against misinformation. *The Conversation*. <https://theconversation.com/meta-is-abandoning-fact-checking-this-doesnt-bode-well-for-the-fight-against-misinformation-246878>
- 30 Rimes, K. A., Shivakumar, S., Ussher, G., Baker, D., Rahman, Q., & West, E. (2019). Psychosocial Factors Associated With Suicide Attempts, Ideation, and Future Risk in Lesbian, Gay, and Bisexual Youth. *Crisis*, 40(2), 83–92. <https://doi.org/10.1027/0227-5910/a000527>
- 31 Mehmet, & Mehmet. (2024, April 23). Annual Review 2023 | ILGA-Europe. ILGA-Europe | Safety, equality and freedom for LGBTI people in Europe & Central Asia. <https://www.ilga-europe.org/report/annual-review-2023/>
- 32 Australian Institute of Health and Welfare. (2023). LGBTQIA+SB people. Suicide & self-harm monitoring. <https://www.aihw.gov.au/suicide-self-harm-monitoring/population-groups/lgbtqiasb-people>
- 33 LGBTIQ+ Health Australia. (2023, July 11). LGBTIQ+ Health Australia (LHA) has called for rigorous evidence over opinions to ensure that trans and gender diverse young people have access to the care they need, at the right time, where they live. [https://www.lgbtiqhealth.org.au/lgbtiq\\_health\\_australia\\_lha\\_has\\_called\\_for\\_rigorous\\_evidence\\_over\\_opinions\\_to\\_ensure\\_that\\_trans\\_and\\_gender\\_diverse\\_young\\_people\\_have\\_access\\_to\\_the\\_care\\_they\\_need\\_atthe\\_right\\_time\\_where\\_they\\_live](https://www.lgbtiqhealth.org.au/lgbtiq_health_australia_lha_has_called_for_rigorous_evidence_over_opinions_to_ensure_that_trans_and_gender_diverse_young_people_have_access_to_the_care_they_need_atthe_right_time_where_they_live)
- 34 Ferlatte, O., Salway, T., Oliffe, J. L., & Trussler, T. (2017). Stigma and suicide among gay and bisexual men living with HIV. *AIDS care*, 29(11), 1346–1350. <https://doi.org/10.1080/09540121.2017.1290762>
- 35 Priest, N. C., Paradies, Y. C., Gunthorpe, W., Cairney, S. J., & Sayers, S. M. (2011). Racism as a determinant of social and emotional wellbeing for Aboriginal Australian youth. *The Medical journal of Australia*, 194(10), 546–550. <https://doi.org/10.5694/j.1326-5377.2011.tb03099.x>
- 36 Allison, F. Cunneen, C., Coombes, L. and Selcuk, A. (2025) "If you don't think racism exists come take a walk with us". *The Call It Out Racism Register 2023-2024*, Sydney: Jumbunna Institute for Indigenous Education and Research, University of Technology Sydney. Available at: Referendum influence on Call It Out racism report | National Justice Project
- 37 Kenny, B., La Sala, L., Gao, C., Cooper, C., McKay, S., Sabo, A., & Robinson, J. (2025). Young People's Behavior, Confidence, and Safety when Communicating Online about Self-Harm and Suicide: Development and Validation of the #Chatsafe Online Safety Scale. *Cyberpsychology, Behavior, and Social Networking*. <https://doi.org/10.1089/cyber.2024.0558>
- 38 Suicide Prevention Australia. (2023). Socio-economic and Environmental Determinants of Suicide.
- 39 Kim, Y. S., & Leventhal, B. (2008). Bullying and suicide. *A review. International journal of adolescent medicine and health*, 20(2), 133–154.
- 40 Holt, M. K., Vivolo-Kantor, A. M., Polanin, J. R., Holland, K. M., DeGue, S., Matjasko, J., & Reid, G. (2015). Bullying and suicidal ideation and behaviors: A meta-analysis. *Pediatrics*, 135(2), e496–e509.
- 41 Perret, L.C., Orri, M., Boivin, M., Ouellet-Morin, I., Denault, A-S., Côté, S.M., Tremblay, R.E., Renaud, J., Turecki, G. & Geoffroy, M-C. (2020). Cybervictimization in adolescence and its association with subsequent suicidal ideation/attempt beyond face-to-face victimization: a longitudinal population-based study. *J. Child Psychol. Psychiatr.* doi: 10.1111/jcpp.13158.
- 42 ReachOut Australia. (n.d.). What is cyberbullying? <https://au.reachout.com/bullying/cyberbullying/what-is-cyberbullying>
- 43 Australian Institute of Health and Welfare. (2023, November 28). Negative online experiences. <https://www.aihw.gov.au/reports/children-youth/negative-online-experiences>
- 44 The NSW Advocate for Children and Young People, The Population Experts, Department of Immigration, ACYP, & Commonwealth Department of Home Affairs. (2024). The experiences of young people from refugee and migrant backgrounds in NSW. <https://522228.fs1.hubspotusercontent-na1.net/hubfs/522228/ACYP%20Refugee%20Migrant%20Report%20Final.pdf>
- 45 American Academy of Pediatrics. (n.d.). The impact of deepfakes, synthetic pornography & virtual child sexual abuse material. Center of Excellence on Social Media and Youth Mental Health. Retrieved May 20, 2025, from <https://www.aap.org/en/patient-care/media-and-children/center-of-excellence-on-social-media-and-youth-mental-health/qa-portal/qa-portal-library/qa-portal-library-questions/the-impact-of-deepfakes-synthetic-pornography--virtual-child-sexual-abuse-material/>
- 46 Jugessur, Y. S. M. F. (2024). Case study (P651) of the little angel MIA, a minor committing suicide due to deepfake AI, CGI, photos manipulation, social media bullying, paparazzi and evil people blackmail. *International Journal of Humanities and Social Science Invention*, 13(6), 25–27. <https://doi.org/10.35629/7722-13062527>
- 47 eSafety Commissioner. (2022, January). Deepfake position statement (Version 2). [https://www.esafety.gov.au/sites/default/files/2022-01/Deepfake-position-statement%20\\_v2.pdf](https://www.esafety.gov.au/sites/default/files/2022-01/Deepfake-position-statement%20_v2.pdf)
- 48 Turnbull, T. (2025, February 8). Woman's deepfake betrayal by friend: "Every moment became porn." <https://www.bbc.com/news/articles/cm21j341m31o>
- 49 Hunt, Gabrielle. (2024, June 12). AI nudes of Victorian students were allegedly shared online. How can schools and parents respond to deepfake porn? *The Guardian*. <https://www.theguardian.com/australia-news/article/2024/jun/12/ai-nudes-of-victorian-students-were-allegedly-shared-online-how-can-schools-and-parents-respond-to-deepfake-porn>
- 50 Center for Countering Digital Hate. (2022). Deadly by design: TikTok pushes harmful content promoting eating disorders and self-harm into young users' feeds. [https://counterhate.com/wp-content/uploads/2022/12/CCDH-Deadly-by-Design\\_120922.pdf](https://counterhate.com/wp-content/uploads/2022/12/CCDH-Deadly-by-Design_120922.pdf)
- 51 Zipfel, S., Giel, K. E., Bulik, C. M., Hay, P., & Schmidt, U. (2015). Anorexia nervosa: Aetiology, assessment, and treatment. *The Lancet Psychiatry*, 2(12), 1099–1111. [https://doi.org/10.1016/S2215-0366\(15\)00356-9](https://doi.org/10.1016/S2215-0366(15)00356-9)
- 52 Fitzsimmons-Craft, E. E., Balantekin, K. N., Graham, A. K., Monterubio, G. E., Sadeh-Sharvit, S., Goel, N. J., Flatt, R. E., Firebaugh, M. L., Eichen, D. M., Taylor, C. B., & Wilfley, D. E. (2021). Technology-assisted delivery of cognitive-behavioral therapy for eating disorders: A systematic review, meta-analysis, and meta-regression. *Journal of Medical Internet Research*, 23(7), e26092. <https://doi.org/10.2196/26092>
- 53 Orygen. (2019, November 13). Young Australian women dying by suicide at increasing rate. <https://www.orygen.org.au/About/News-And-Events/2019/Young-Australian-women-dying-by-suicide-at-increas>
- 54 Bearman, S. K., Presnell, K., Martinez, E., & Stice, E. (2006). The skinny on body dissatisfaction: A longitudinal study of adolescent girls and boys. *Journal of Youth and Adolescence*, 35(2), 217–229.
- 55 Hinduja, S., & Patchin, J. W. (2010). Bullying, cyberbullying, and suicide. *Archives of Suicide Research*, 14(3), 206–221.
- 56 Fuller, K. (2021, April 28). The relationship between sexual assault, self-harm, and trauma. *Psychology Today*. <https://www.psychologytoday.com/au/blog/happiness-is-state-mind/202104/the-relationship-between-sexual-assault-self-harm-and-trauma>
- 57 Suicide Prevention Australia (2023). Socio-economic and environmental determinants of suicide. Background Paper.
- 58 Tsai, A. C., Lucas, M., & Kawachi, I. (2015). Association Between Social Integration and Suicide Among Women in the United States. *JAMA psychiatry*, 72(10), 987–993.
- 59 Fitzpatrick, S. J., Handley, T., Powell, N., Read, D., Inder, K. J., Perkins, D., & Brew, B. K. (2021). Suicide in rural Australia: A retrospective study of mental health problems, health-seeking and service utilisation. *PloS one*, 16(7), e0245271.
- 60 Common Sense Media. (2024, January 25). Parents' ultimate guide to AI companions and relationships. <https://www.common SenseMedia.org/articles/parents-ultimate-guide-to-ai-companions-and-relationships>

- 61 Common Sense Media. (2024, January 25). Parents' ultimate guide to AI companions and relationships. <https://www.common sense media.org/articles/parents-ultimate-guide-to-ai-companions-and-relationships>
- 62 Walther, C. C. (2024, September 9). Can we use AI to feel less lonely? Forbes. <https://www.forbes.com/sites/corneliawalther/2024/09/09/can-we-use-ai-to-feel-less-lonely/>
- 63 Morse, J. (2024, October 23). Character.AI sued after Florida teen's death; chatbot allegedly encouraged suicide. NBC News. <https://www.nbcnews.com/tech/characterai-lawsuit-florida-teen-death-rcna176791>
- 64 Suicide Prevention Australia (2023). Socio-economic and environmental determinants of suicide. Background Paper.
- 65 Australian Institute of Health and Welfare. (2023). Income variability and uncertainty. Suicide & self-harm monitoring. <https://www.aihw.gov.au/suicide-self-harm-monitoring/risk-factors/social-economic-factors/income-variability-uncertainty>
- 66 Scamwatch. (2025, March 5). Scam statistics. Scamwatch. <https://www.scamwatch.gov.au/research-and-resources/scam-statistics>
- 67 The Guardian. (2024, November 7). Victorian teenager's suicide after sextortion scam prompts call for improved education. The Guardian. <https://www.theguardian.com/australia-news/article/2024/jul/22/victorian-teenagers-suicide-after-sextortion-scam-prompts-call-for-improved-education>
- 68 Duffin, P. (2024, April 7). Sextortion pushed a NSW teen to suicide, cops traced culprits to Nigeria. The Sydney Morning Herald. <https://www.smh.com.au/national/nsw/sextortion-pushed-a-nsw-teen-to-suicide-cops-traced-culprits-to-nigeria-20240405-p5fhnp.html>
- 69 Aslett, D., Mekonnen, G. T., Lin, L. S., & Zecevic, M. (n.d.). The dangers of voice cloning and how to combat it. The Conversation. <https://theconversation.com/the-dangers-of-voice-cloning-and-how-to-combat-it-239926#:~:text=In%202022%2C%20almost%20240%2C000%20Australians,financial%20loss%20of%20%24A568%20million.>
- 70 Campbell, D., & McCutchan, E. (2023, August 3). The Robodebt scheme resulted in significant harm, including suicides. But the 2,030 deaths figure spread by some is not supported by the data. ABC News. <https://www.abc.net.au/news/2023-08-04/fact-check-2030-robodebt-deaths/102684206>
- 71 RNZ News. (2023, July 8). Robodebt: Illegal Australian welfare hunt drove people to despair. RNZ. <https://www.rnz.co.nz/news/world/493404/robodebt-illegal-australian-welfare-hunt-drove-people-to-despair>
- 72 Hamilton, E. August 09, 2023. Robodebt Royal Commission Findings. Law Society Journal, <https://lsj.com.au/articles/crude-cruel-and-unlawful-robodebt-royal-commission-findings/>
- 73 Australia's Robodebt scheme: A tragic case of public policy failure. (2023, July 26). Blavatnik School of Government. <https://www.bsg.ox.ac.uk/blog/australias-robodebt-scheme-tragic-case-public-policy-failure>
- 74 Wei, M., MD JD. (2024, November 4). New research suggests learning about AI helps us trust AI a little more. Psychology Today. <https://www.psychologytoday.com/gb/blog/urban-survival/202411/do-we-trust-ai-to-help-make-decisions-for-mental-health>
- 75 Reset Australia. (2024). Accountability & safety requirements: A framework for the Australian context. Reset Australia. <https://au.reset.tech/uploads/Accountability-&-Safety-requirements-0424-V2.pdf>
- 76 Chhabra, J., Pilkington, V., & Seidler, Z. (n.d.). Banning young people from social media sounds like a silver bullet. Global evidence suggests otherwise. The Conversation. <https://theconversation.com/banning-young-people-from-social-media-sounds-like-a-silver-bullet-global-evidence-suggests-otherwise-256587>
- 77 Ministers for the Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts. (2024, November 14). Ministers for the Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts. Ministers for the Department of Infrastructure. <https://minister.infrastructure.gov.au/rowland/media-release/new-duty-care-obligations-platforms-will-keep-australians-safer-online>
- 78 Perrino, J. (2022, July 26). Using 'safety by design' to address online harms. Brookings. <https://www.brookings.edu/articles/using-safety-by-design-to-address-online-harms/>
- 79 Given, L. M. (n.d.). Australia will impose a 'digital duty of care' on tech companies to reduce online harm. It's a good idea – if it can be enforced. The Conversation. <https://theconversation.com/australia-will-impose-a-digital-duty-of-care-on-tech-companies-to-reduce-online-harm-its-a-good-idea-if-it-can-be-enforced-243682#:~:text=The%20digital%20duty%20of%20care,to%20proactively%20identify%20harmful%20content.>
- 80 Suicide Prevention Australia. (2022). Submission: Online safety rules. <https://www.suicidepreventionaustralia.org/wp-content/uploads/2022/09/Suicide-Prevention-Australia-Submission-Online-Safety-Rules.pdf>
- 81 Reset Tech. April 2024. A duty of care in Australia's Online Safety Act Policy briefing. Accessed at: [Duty-of-Care-Report-Reset.Tech.pdf](https://www.reset.tech/uploads/Duty-of-Care-Report-Reset.Tech.pdf)
- 82 Dittmar, L. (2024, November 8). What does transparency really mean in the context of AI governance? OCEG. <https://www.oceg.org/what-does-transparency-really-mean-in-the-context-of-ai-governance/>
- 83 TechTarget. (2022, November 10). algorithmic transparency. Search Enterprise AI. <https://www.techtarget.com/searchenterpriseai/definition/algorithmic-transparency#:~:text=Algorithmic%20transparency%20is%20openness%20about,for%2C%20process%20and%20deliver%20information.>
- 84 Unraveling the psychology of trust in artificial intelligence. <https://www.immersivelabs.com/resources/blog/unraveling-the-psychology-of-trust-in-artificial-intelligence>
- 85 Laidler, John. 2019. The Harvard Gazette. Harvard professor says surveillance capitalism is undermining democracy. Accessed at: [Harvard professor says surveillance capitalism is undermining democracy — Harvard Gazette](https://www.harvard.edu/gazette/2019/09/harvard-professor-says-surveillance-capitalism-is-undermining-democracy)
- 86 Blackman, R., & Ammanath, B. (2022, June 20). Building Transparency into AI Projects. Harvard Business Review. <https://hbr.org/2022/06/building-transparency-into-ai-projects>
- 87 Reset.Tech Australia. (2024). Regulating for transparency: Transparency reports in Australia. In Reset.Tech Australia. <https://au.reset.tech/uploads/Regulating-for-Transparency-Reset.Tech.pdf>
- 88 Orygen. (n.d.). Co-designing with young people: The fundamentals. <https://www.orygen.org.au/Training/Resources/Service-knowledge-and-development/Guidelines/Co-designing-with-young-people-The-fundamentals/Orygen-Co-designing-with-YP-the-fundamentals?ext=>



Suicide Prevention  
Australia

## Acknowledgements Statement

Suicide Prevention Australia acknowledges the unique and important understanding provided by people with lived and living experience. This knowledge and insight is critical in all aspects of suicide prevention policy, practice and research. As described above, advice from individuals with lived experience helped guide the analysis and recommendations outlined in this paper.

As the national peak body for suicide prevention, our members are central to all that we do. Advice from our members, including the largest and many of the smallest organisations working in suicide prevention, as well as practitioners, researchers and community leaders is key to the development of our papers. Suicide Prevention Australia thanks all involved in the development of this paper.

**If you or someone you know require 24/7 crisis support, please contact:**

**Lifeline: 13 11 14**

[www.lifeline.org.au](http://www.lifeline.org.au)

**Suicide Call Back Service: 1300 659 467**

[www.suicidcallbackservice.org.au](http://www.suicidcallbackservice.org.au)

**For general enquiries**

02 9262 1130 | [policy@suicidepreventionaust.org](mailto:policy@suicidepreventionaust.org) | [www.suicidepreventionaust.org](http://www.suicidepreventionaust.org)