Review

Social Media and Youth Mental Health: Scoping Review of Platform and Policy Recommendations

Jasleen Chhabra[1], MBBSMD, PhD; Vita Pilkington[1], BA (Hons); Ruben Benakovic[1], BA (Hons); Michael James Wilson[1], BA (Hons); Louise La Sala[1], PhD; Zac Seidler[2], PhD

[1]Centre of Youth Mental Health, Orygen Youth Health, Footscray, Australia

[2]Movember Foundation, East Melbourne, Australia

Corresponding Author:
Jasleen Chhabra, MBBSMD, PhD
Centre of Youth Mental Health
Orygen Youth Health
35 Poplar Road, Parkville
Footscray, 3052
Australia
Phone: 61 452642643
Email: jasleen.chhabra@orygen.org.au

Abstract

Background: High rates of social media use and mental ill-health among young people have drawn significant public, policy, and research concern. Rapid technological advancements and changes in platform design have outpaced our understanding of the health effects of social media and hampered timely evidence-based regulatory responses. While a proliferation of recommendations to social media companies and governments has been published, a comprehensive summary of recommendations for protecting young people's mental health and digital safety does not yet exist.

Objective: This scoping review synthesized published recommendations for social media companies and governments in relation to young people's (aged 12-25 years) mental health. A qualitative approach was used to undertake inductive content analysis, where recommendations were grouped under conceptually similar themes.

Methods: We searched academic (PubMed, Scopus, and PsycINFO) and nonacademic (Overton and Google) databases for relevant documents. Eligible documents provided recommendations to regulators and social media companies that pertained to social media, young people, and mental health. This review excluded recommendations for young people, caregivers, educators, or clinicians surrounding strategies for managing individual social media use; instead, the recommendations emphasized the regulation or design of social media products and practices of social media companies. Peer-reviewed and gray literature from selected Western contexts (Australia, Canada, the United Kingdom, and the United States) were relevant for inclusion. Documents were published between January 2020 and September 2024.

Results: Of the identified 4980 unique reports, 120 (2.41%) progressed to full-text screening, and 70 (1.41%) met the inclusion criteria. Five interrelated themes were identified: (1) legislating and overseeing accountability, (2) transparency, (3) collaboration, (4) safety by design, and (5) restricting young people's access to social media.

Conclusions: This review emphasizes the need for multipronged approaches to address the rapidly increasing presence and reach of social media platforms in the lives of young people. These recommendations provide practical and tangible paths forward for governments and industry, backed by expert organizations in youth mental health and technology regulation at a time when expert-informed guidance is sorely needed. Rigorous evaluation of the proposed recommendations is needed while continuing to build on the emerging peer-reviewed evidence base that should form the foundation of policy and regulatory changes.

XSL•FO
RenderX

Introduction

Background

The impact of social media on young people's mental health has emerged as a pressing and global issue of broad public interest, captivating the attention of policy makers, caregivers, educators, and researchers. In Western nations, high rates of social media use among young people indicate the extent to which social media has become a central component in their daily lives, development, and social experiences. In Australia, 98% of young people regularly use at least one social media platform [1]. Similar statistics are observed in other countries, such as the United Kingdom, where an estimated 92% of young people are active on social media by the age of 12 years [2]; the United States, where 95% of those aged 13 to 17 years report using social media [3]; and Canada, where 91% of those aged 15 to 24 years use social media [4].

The definition of "young person" varies across cultural and societal contexts. Following established consensus, we define "young people" as individuals aged 12 to 25 years [5]. This is a time of significant transition, where individuals negotiate and develop their identities, relationships, goals, and values, all of which interact with mental health and well-being [5]. Social media undoubtedly plays a key role in young people's daily lives as they increasingly use it to shape their identities and forge new connections [6,7]. Young people consistently report social and personal benefits of social media. Social media can be instrumental in maintaining their offline relationships, expanding their social networks, and fostering a sense of belonging [7,8]. Social media also offers opportunities for identity exploration, providing connection among groups that experience marginalization, such as lesbian, gay, bisexual, transgender, and queer (LGBTQ+) communities [9,10], and can facilitate access to information that may not be available or accessed through other means [2,11].

Young People's Mental Health and Social Media Content

The social and personal affordances of social media may translate into various mental health benefits. Mental health is a broad, multifaceted construct that encompasses an individual's emotional, psychological, and social well-being [12]. The social aspects of these web-based platforms, such as the ability to message others and share content, have been linked to greater well-being, reduced depression, increased positive emotions, and decreased loneliness [13-17]. Being connected to support groups can also help young people feel validated and understood and improve their ability to manage their mental ill-health (an umbrella term that encompasses a continuum from the most commonly occurring to the most severe and disabling mental health problems [18]) or suicidal thoughts and behaviors [19]. The often entertaining and motivational nature of certain content can also provide an outlet for young people to relax, have fun, and learn new information or skills, thereby increasing their emotional and psychological well-being [20]. However, such benefits are not universally recognized by others in young people's lives. In recent Australian surveys of 631 parents and 921 young people (aged 16-25 years), parents ranked social media as the leading concern in relation to their children's well-being, whereas young people ranked social media as the 24th most pressing concern regarding their well-being [21,22]. This exemplifies a generational disconnect where young people often cite important benefits of social media, whereas caregivers and other adults focus to a far greater extent on its potential harms.

Exposure to harmful content and interactions on social media has increasingly been highlighted as a significant risk factor for young people's mental health [2,23]. For example, exposure to content related to self-harm and suicidality has been associated with higher rates of depression and anxiety [24] and, in some cases, has been linked to suicide attempts in young people [25]. Exposure to disordered eating content, particularly among adolescent girls [26,27], has been associated with poor self-esteem [28], body image concerns, and eating disorders [29-31]. Cyberbullying, a prevalent form of harassment on social media that is reported by 44% of Australian young people [32], is associated with depression, anxiety, and suicidal ideation [2,24,33].

Given their capacity for social connection, social media platforms can become sites for predatory behaviors and interactions, including distribution of child sexual abuse material and adults seeking to sexually exploit children [34,35]. Exposure to such content can have severe implications for young people's emotional and psychological well-being, including potentially leading to posttraumatic stress disorder [36] and acute stress disorder [37]. In addition, social media platforms can be used to amplify misinformation (false, misleading, or inaccurate information spread without deliberate intent to deceive) and disinformation (false information deliberately spread to mislead or deceive, which can include manipulated facts, narratives, and propaganda) [38]. Both misinformation and disinformation can jeopardize young people's understanding of reality, resulting in confusion, fear, and paranoia, and perpetuate harmful attitudes and beliefs [39]. For example, the spread of health-related misinformation (eg, overstating the risks associated with COVID-19 vaccination [38]) has been associated with poor vaccine uptake and mental ill-health in young people (eg, depression, social anxiety, and loneliness [40]). Indeed, young people's exposure to extremism and hate speech on social media has been linked to the spread of radical ideologies and extremist views [41,42]. Exposure to such content can impact mental health by creating a sense of alienation from peers and communities, resulting in isolation, anxiety, and even violence [43], necessitating platform features for inhibiting this harmful content.

Social media companies use a range of strategies for moderating posts to limit the proliferation of harmful content on their platforms, including human-facilitated and automated methods (eg, artificial intelligence [AI] tools [44]). However, there are concerns that users can easily circumvent moderation practices given that moderation tools can fail to detect harmful posts and have been outpaced by the increasing use of AI and bots on social media [45-47]. These limitations can result in the proliferation of harmful content that may pose direct risks to the mental health of young people.

Young People's Mental Health and Social Media Design

In addition to harmful posts on social media platforms, some platform design features may inherently impact young people's mental health. Rapid advancements in the development and adoption of content recommender systems (a platform design that organizes, prioritizes, and promotes content into users' feeds based on user actions, preferences, and demographics [48]) have also raised concerns [49]. Recommender systems and highly engaging design features such as "endless scrolling" (a listing page design approach that loads content continuously as the user scrolls down [50]) are used by social media companies to keep users engaged with their platforms for extended periods [51]. Moreover, recent changes to recommender systems have resulted in deprioritizing content from family, friends, and the accounts that users follow in favor of algorithmically curated content [52]. These curated, highly personalized feeds, which are shaped by user data, are designed to keep users on platforms as long as possible [53]. Some suggest that this design feature can amplify young people's exposure to extreme viewpoints (eg, misogyny, homophobia, racism, or eating disorder content [54]), inappropriate advertising material (eg, promotion of cosmetic procedures and diet supplements), and disinformation [55]. Importantly, engaging app features and curated content have been associated with sleep disturbance, compulsive behaviors (eg, continuously refreshing social media apps), depression, and anxiety, resulting in overall poor mental health and well-being [56,57]. The role of recommender systems in young people's routine exposure to harmful content is particularly concerning given evidence of rapid exposure to such content despite minimal user engagement. For example, a report [58] found that faux YouTube accounts designed to reflect the demographic characteristics of male users aged 12 to 20 years were frequently exposed to harmful content that promoted restrictive masculine norms and antifeminist rhetoric even in the absence of user engagement, such as watching videos, liking content, or subscribing to channels. This highlights that, even with minimal user input, and in the absence of effective moderation, potentially harmful content can be rapidly pushed to young people's feeds.

Legislation by governments and regulators such as the eSafety Commissioner in Australia [59], Ofcom in the United Kingdom [60], the Federal Communications Commission and Federal Trade Commission in the United States [61,62], and the Canadian Radio-television and Telecommunications Commission in Canada [63] regulate social media platforms, including their design features, to minimize negative mental health impacts on young people [64]. However, rapid advancements in technology have led to an "arms race" [65] between regulators and social media companies, where the pace of technological advancements and sophistication of algorithms have eclipsed the capacity for timely updates to safety and regulation [64,66]. Moreover, the lack of specificity in existing legislation (eg, allowing legal but likely harmful content such as targeted advertisements [67]) and the lack of international consistency in legislation [64] further complicate efforts to minimize negative mental health impacts on young people. Consequently, there are growing concerns that the largely autonomous governance strategies adopted by social media platforms, coupled with the limitations of current legislations, are insufficient to mitigate the risks to young people's mental health. As a result, there are calls for stricter regulation, legislative reform, and greater oversight of social media companies' practices [68,69]. One approach to mitigate negative impacts of social media on young people that is being pursued in some Western countries is restricting social media access. For example, the Australian government has recently legislated restriction of young people's (aged <16 years) access to social media from 2025 [70]. Similar measures are being discussed in both the United States and the United Kingdom, with the United Kingdom considering the potential restriction of young adults' access to social media [71].

This Review

Growing public concern and recent developments in legislation have led to a proliferation in reports and policy documents offering recommendations to governments and social media companies regarding strategies for more effective social media regulation, design, and practices, with an emphasis on safeguarding young people's mental health. While this is a promising development, the academic literature presently lacks a comprehensive synthesis of recommendations. Therefore, this scoping review aimed to consolidate existing recommendations regarding how social media can be effectively designed and regulated to prioritize the protection of young people's mental health and minimize harm. This review synthesized available recommendations directed toward governments, regulators, and social media companies using thematic analysis to group them into categories reflecting similar aims or principles.

Methods

Search Strategy

This review and reporting methodology followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) statement [72]. The search strategy was prepared in consultation with a university librarian. Included documents were identified through 2 sources (Overton and Google) using the following search terms: "Social Media" AND "Young People" AND "Mental Health" AND "Recommendations" OR "Guidelines." Given the focus on practical, policy-oriented recommendations in this review, the search strategy primarily drew on gray literature sources. Although peer-reviewed literature was eligible for inclusion (and was also considered and trialed when developing the search strategy but yielded few relevant documents), it was deemed more appropriate to prioritize gray literature documents that offered tangible recommendations to governments, regulators, and members of the social media industry, such as policy briefs and reports from nongovernment agencies and advocacy bodies. Overton was chosen as it is the world's largest repository of policy documents, guidelines, think tank publications, and working papers [73] and, therefore, was particularly relevant to our focus on documents published primarily outside of academic journals. Broad search terms were used to conduct this search as both Overton and Google automatically interpret the terms and use synonyms to capture

relevant documents. Additional documents included were extracted from the reference lists of the identified documents or those recommended by expert colleagues. To ensure a comprehensive search of the literature, we reran our search strategy in the following academic databases—PubMed, Scopus, and PsycINFO—but no additional relevant results were identified.

Selection Criteria

To be included in this study, documents were required to (1) focus on young people between the ages of 12 and 25 years; (2) present recommendations for social media companies, governments, or regulators that promote safe social media products and services for young people; (3) be peer-reviewed published journal articles (original research), case studies, or gray literature (eg, policy briefs and reports); (4) be published in Australia, Canada, the United Kingdom, and the United States or published by global organizations (eg, Amnesty International or the World Health Organization); (5) be published between January 2020 and September 2024 (due to the proliferation of social media platforms and their users after the COVID-19 pandemic [74]); and (6) be published in English.

Data Screening

Search outputs were imported to the Covidence systematic review software (Veritas Health Innovation) [75]. The first 2 authors independently screened all titles and abstracts against the inclusion criteria. Results were compared, with a disagreement rate of <2% recorded (Gwet AC1=0.97; Cohen κ=0.54). Discrepancies were discussed until consensus was reached between the authors. For the second stage of the review,

dual screening was conducted on 25% of the full-text documents. Having achieved 0% disagreement rate, single-author screening proceeded thereafter, which was then checked for accuracy by the second and third author.

Data Extraction and Synthesis

The extracted data included title, authors, organization, year, country, type of document, target outcomes, and target audience. Information such as recommendations to social media companies (eg, improved platform design) and government bodies (eg, strengthening existing legislations) was extracted. To collate and synthesize the recommendations provided, all documents were exported to NVivo (version 14; QSR International) [76]. The first and second authors analyzed the included recommendations, grouping them into preliminary themes and subthemes to capture conceptually similar ideas. The preliminary themes were then reviewed by the wider coauthor team, and through a collaborative process, the thematic structure was refined to ensure conceptual clarity, coherence, and alignment with the study aims.

Results

Overview

The search strategy returned 6966 reports, of which 1911 (27.43%) were duplicates. In the initial screening stage, 4980 reports were assessed for relevance. Following the exclusion of 97.59% (4860/4980) of nonrelevant documents, 120 reports progressed to full-text screening, leading to the inclusion of 70 (58%) reports in this scoping review (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram.
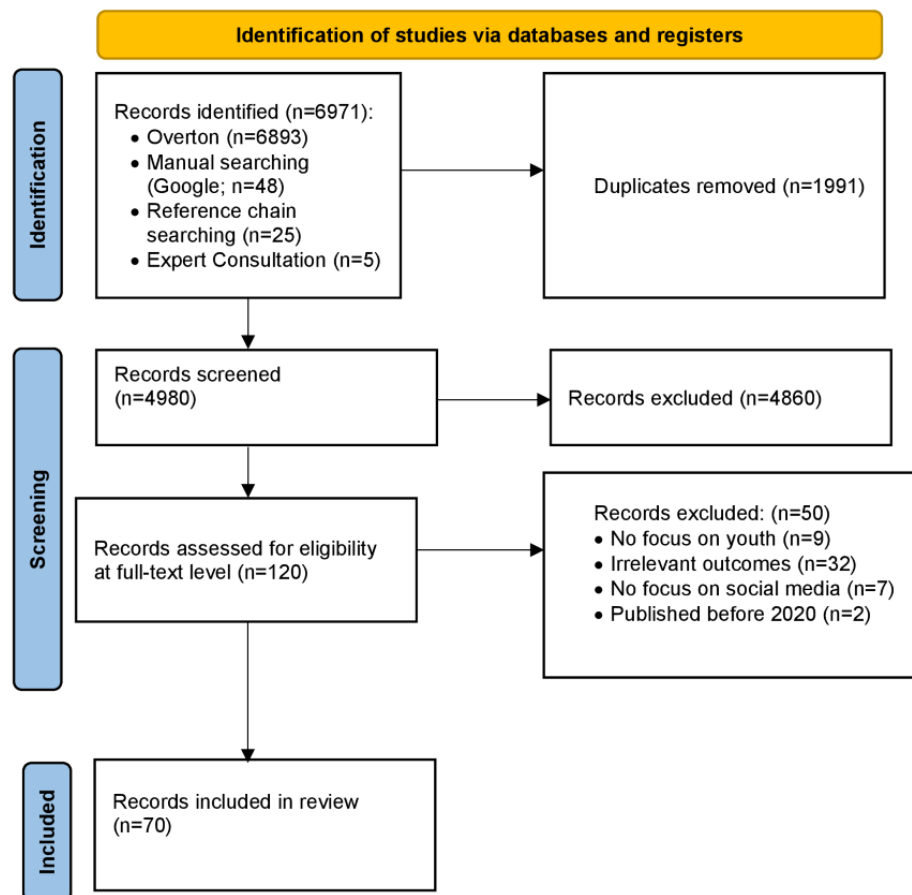
Report Characteristics

Of the 70 included reports, 17 (24%) provided recommendations to regulators or governments, 14 (20%) provided recommendations to social media companies, and 39 (56%) provided recommendations to both regulators or governments and social media companies. The reports sought to target the following outcomes for young people: mental health and well-being related to social media use (35/70, 50%), including exposure to content related to self-harm and suicidality (7/70, 10%); exposure to misinformation and disinformation (16/70, 23%); exposure to discrimination, hate crime, and hate speech (6/70, 9%), including extremist content (9/70, 13%), misogynistic content (5/70, 7%), and cyberbullying (3/70, 4%); exposure to content that increases risk of body image concerns and disordered eating (8/70, 11%); and exposure to pornography and digital child sexual abuse and exploitation (9/70, 13%). Most reports were published in Australia (22/70, 31%), the

United Kingdom (18/70, 26%), and the United States (17/70, 24%). The characteristics of the included reports are shown in Multimedia Appendix 1 [3,10,31,33,39,41,52,54,56,58,65,69, 77-134].
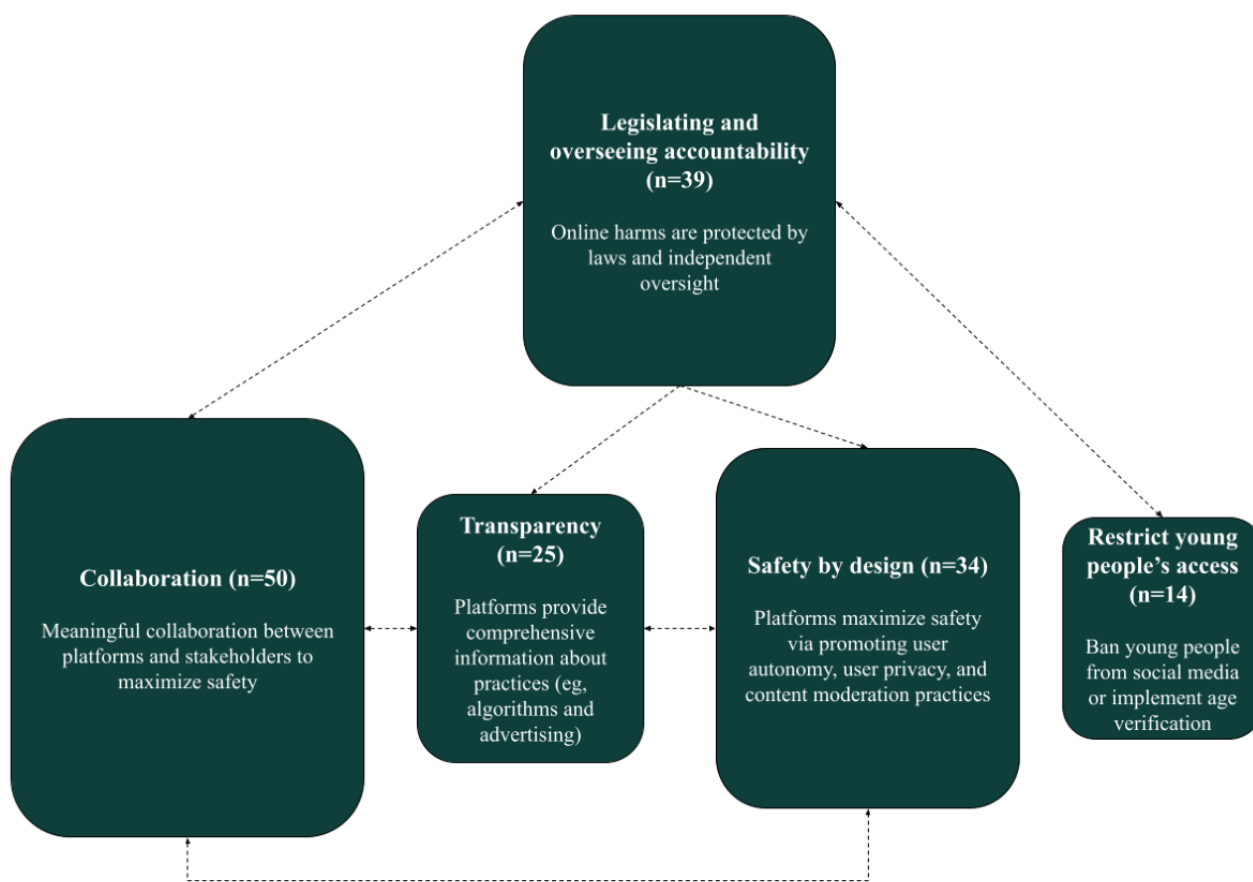
Synthesis of Evidence

The content analysis resulted in the identification of 5 core interrelated themes, as shown in Figure 2. A PRISMA-ScR checklist was also completed to demonstrate the methodological rigor of this evidence (Multimedia Appendix 2). In total, 4 of the 5 themes centered on strategies for social media companies and regulators to facilitate the provision of safer products to young people, whereas 1 theme centered on restricting young people's access to social media altogether.

The 5 themes are described briefly in Table 1 and in further detail in the following sections, with illustrative quotes throughout.

Figure 2. Thematic map of the 5 core interrelated themes of social media recommendations. Note: The size of each box corresponds to the frequency at which each theme was present in the included reports, with larger size indicating that the theme was present in a higher number of reports.

Table 1. Description of the themes and subthemes.

Theme	Subthemes
Theme 1: legislating and overseeing accountability (n=39)—recommendations regarding the need for governments and independent regulatory bodies to hold greater levels of responsibility and accountability to ensure that (1) users are not harmed; (2) when users are harmed, the harms are addressed appropriately; and (3) failure to address harmful content leads to penalties for social media companies	• Creating or reforming legislation—legislative changes to facilitate ensuring that social media companies are accountable in protecting user safety • Expanding the definition of online harms—broadening the current definitions of online harms, including incorporating harms associated with body image and technology-facilitated gender-based violence • Empowering and equipping regulators—strategies for ensuring that regulators are empowered and equipped to hold social media companies accountable
Theme 2: transparency (n=25)—recommendations on the ways in which social media companies can facilitate access to clear information about their products and practices (eg, with users, governments, and researchers)	• Advertising and algorithmic transparency—social media platforms should provide users and others (eg, regulators and researchers) with clear information regarding promotional material on users' feeds and how users' algorithms are shaped • Transparency reports—social media platforms should release regular prescriptive transparency reports that offer regulators and the public clear and detailed information about content moderation policies and practices, accuracy and error rates for automated review, and number of "take-down" orders, among other things
Theme 3: collaboration (n=50)—recommendations suggesting that social media companies collaborate with relevant stakeholders (eg, parents, caregivers, regulators, and researchers) at all stages of service design and delivery	—[a]
Theme 4: safety by design (n=34)—recommendations on how social media companies can design, develop, and deploy products that are safe to individuals using their services or platforms; includes handling user data and ensuring that potential harms to users are minimized, identified, and addressed	• Privacy—recommendations to enhance user privacy and data protection legislations to limit the collection and use of user data by social media companies • Content moderation—social media companies' roles in determining content that is appropriate and safe to appear on their platforms and actions taken to prevent harmful digital content from appearing and proliferating on these platforms and remove it • Autonomy—strategies adopted by social media companies (including those responsible for designing, developing, and delivering services) to increase platform users' decision-making about their (the users') experiences on social media platforms
Theme 5: restricting young people's access to social media (n=14)—recommendations regarding implementing age verification measures and banning young people from accessing social media platforms or services until they reach specified ages	—

[a]Not applicable.

Theme 1: Legislating and Overseeing Accountability

Overview

The first theme, legislating and overseeing accountability, encompasses recommendations regarding the need for governments and independent regulatory bodies to hold greater levels of responsibility and capacity to effectively oversee the practices of social media companies. The accountability is to ensure that (1) users are not harmed; (2) when users are harmed, the harms are addressed appropriately; and (3) failure to address harmful content leads to penalties for social media companies. As such, these recommendations aim to enshrine user safety and the accountability of social media companies in law and empower regulators to enforce this with consequences for breaches of liability. Theme 1 contained 3 subthemes, which are discussed in the following sections.

Theme 1a: Creating or Reforming Legislation

The subtheme *creating or reforming legislation* appeared in 20% (14/70) of the reports and referred to legislative changes required to ensure that social media companies are held accountable for protecting young users' mental health. Many recommendations centered on making social media companies

liable for allowing dissemination of harmful user-generated content on their platforms [77]. This included imposing penalties "sufficient to realise the goal of deterrence" [78]. Some reports (8/14, 57%) referred broadly to "harmful material" posted on social media platforms, whereas others (6/14, 43%) specifically discussed the need to impose liabilities regarding content that encourages extreme dieting, eating disorders, self-harm [79], and violent crimes [78] (quote from Canadian Commission on Democratic Expression [77]):

> *Legislation can and should leave open the possibility of platforms being held liable for violating their duty to act responsibly in relation to the algorithmic curation and amplification of certain kinds of content.*

A number of reports (2/14, 14%) recommended changes to Section 230(c)(1-2) of the 1996 Communications Decency Act in the United States [78,80], which currently offers social media companies immunity from responsibility for user-generated content on their platforms. With the amendment of this law, regulators would be able to hold owners of platforms such as TikTok and Instagram (eg, Meta and ByteDance) accountable for harmful posts (eg, related to self-harm or hate speech) posted and shared by users. Implementing such legislative changes,

with associated penalties for noncompliance, would likely encourage platforms to enact more effective self-regulation [80] (quote from Office of the New York State Attorney General Letitia James [78]):

> *Instead of simply being able to assert protection under Section 230, a defendant company has the initial burden of establishing that its policies and practices were reasonably designed to address unlawful content.... This would help establish a baseline of content moderation across online platforms, helping to ensure a safer experience for all users.*

Theme 1b: Expanding Definitions of Online Harms in Legislation

The subtheme *expanding definitions of online harms* referred to recommendations in 13% (9/70) of the reports regarding broadening current definitions of online harms in relevant legislation (eg, Australia's Online Safety Act). Amending this definition was suggested to ensure consistent and comprehensive prevention and response measures regarding content on social media platforms deemed likely to cause harm to users [81].

Several specific online harms were noted in the included reports. For example, the Women and Equalities Committee [82] called for inclusion of appearance-based bullying in UK legislation pertaining to online harms; the Butterfly Foundation and Daniel [81] recommended inclusion of harms associated with body image in the Australian Online Safety Act; and the Royal Society for the Encouragement of Arts, Manufactures, and Commerce [83] suggested that the Online Safety Act include harms caused by misinformation in its remit. Furthermore, one report that focused on extremism and hate speech on TikTok suggested that the platform should offer a clear definition of extremism in their terms of service, arguing that effective restriction of harmful material hinges on clear definitions of harm [41] (quote from The Butterfly Foundation and Daniel [81]):

> *The Online Safety Act 2021 (Cth) should be modified so that members of the Australian public can report to the eSafety Commissioner when they see material that could negatively affect their body image.*

The Institute for Strategic Dialogue (ISD) [39] reflected on the need for legislation to adapt to emerging virtual reality spaces, including what constitutes crimes within the Metaverse. The ISD [39] argued that uncertainty about how offenses such as sexual assaults occur within these emerging digital spaces warrants the need to review existing criminal codes, define criminal behavior in virtual spaces, and ensure that legislation exists to prosecute such offenses (quote from Dorn et al [39]):

> *It should be argued that the immersiveness of the Metaverse can result in the very real (emotional) experience of rape, reiterating questions about the requirements of existing laws to quality for the legal terms of rape.*

Theme 1c: Empowering and Equipping Regulators

The subtheme *empowering and equipping regulators* appeared in 43% (30/70) of the reports and referred to the need to implement strategies that ensure that independent regulators are equipped to hold social media companies accountable for ensuring young users' safety and addressing any harms encountered when young people use social media.

Specific recommendations included conducting routine independent audits [84], establishing independent regulators (including data protection regulators [85]), and ensuring that regulators have power to impose penalties when social media companies fail to comply with relevant rules and regulations [65,77]. Within these reports, the authors highlighted the importance of external oversight in light of the current norm of self-regulation of practices and policies by social media companies [84-86] (quote from Amnesty International [85]):

> *Stronger laws and regulation on data protection and algorithmic amplification of content on social media, and effective enforcement of such laws and regulation, are needed in order to keep children safe from the harvesting and exploitation of their personal data for profit.*

Several reports (7/30, 23%) recommended establishing new independent regulators focused on auditing, investigating, and suggesting reforms to the policies and practices of social media platforms and their enforcement [77,78]. The authors noted the need for such regulators to have information-gathering powers and access data held by social media companies for oversight purposes [65,77]. This also requires regulators to be appropriately informed and equipped to deal with the complex, ever-changing nature of digital technologies and behavior [87] (quote from Reset Australia [65]):

> *To meaningfully drive change, regulations need to be enforceable, and regulators must be empowered and resourced.*

Theme 2: Transparency

Overview

The second theme, *transparency*, encompassed recommendations regarding ways in which social media companies can increase the clarity of information regarding their products and practices. This includes communicating with users, governments and regulators, and researchers. This theme contained two subthemes: (1) *advertising and algorithmic transparency*, which related to transparency about what users see on their platforms or feeds; and (2) *transparency reports*, which related to the need for consistent and prescriptive official reporting of social media companies' practices and policies, particularly regarding content moderation.

Theme 2a: Advertising and Algorithmic Transparency

Calls for advertising and algorithmic transparency appeared in 23% (16/70) of the reports. These recommendations emphasized the importance of offering greater transparency about what users are exposed to when using social media platforms. These recommendations argue that social media companies should provide users with clearer information regarding promotional material on users' feeds and how algorithms are shaped.

Development of advertising repositories, or openly searchable databases of all advertisements presented on social media platforms, was suggested in several reports (5/16, 31%) [65,69,87]. It was recommended that these repositories include clear details about the advertiser, advertisement content, targeting parameters (ie, specific demographics of the users being targeted), demographics of those exposed to advertisements, and the advertising spending and length of campaigns [77,84,87]. Several reports (3/16, 19%) noted a need for greater transparency surrounding political advertising in particular [77,84,87]. For example, during the US 2024 elections, several reports (3/16, 19%) suggested that disinformation on platforms such as X (formerly known as Twitter) and Facebook heavily influenced individuals' views of candidates and their policies on the economy, immigration, and crime [135] (quote from Canadian Commission on Democratic Expression [77]):

> *Advertising transparency will aid in the identification of discriminatory and biased advertising practices...Even advertisers would benefit from transparency that provides them with more data about how platforms manage their advertisements, details about views and clicks on the advertisement and the influence of bots rather than humans, than is currently made available to them.*

Recommendations also advocated for transparency regarding algorithms, where social media companies should articulate to users how their algorithms (or "feeds") are shaped [81,88]. The authors noted that social media algorithms operate "opaquely, with little transparency regarding how content is prioritised and served to users" [52] (quote from the Jed Foundation [89]):

> *[We recommend] Supporting federal regulation designed to maximise protective factors, including...promoting transparency in algorithms across social media platforms, ensuring that young social media users gain insight into the factors shaping their online experiences.*

Recommendations also pertained to social media influencers and celebrities. These included calls for these figures to be made subject to stricter regulation, ensuring that any promotional, paid, or incentivized posts are clearly signposted to users [31]. It is important to acknowledge that, currently, promotional or paid posts are often labeled as "paid." However, the rapid increase in the number of social media influencers, their followers (young people), and the promotional materials presents new challenges for advertising regulators, requiring stricter disclosure and transparency standards [82].

Theme 2b: Transparency Reports

A total of 23% (16/70) of the reports called for transparency reports to be released regularly by social media companies, offering clear and publicly accessible information about social media companies' policies and practices, particularly surrounding content moderation and removal. It was acknowledged that transparency reports have begun to be released by key platforms, although the authors noted the need for clear rules regarding what should be included in these reports to facilitate *meaningful* transparency [84-86,90]. In particular,

the Office of the New York State Attorney General Letitia James [78] and Reset Australia [84] offered a number of tangible recommendations regarding what should be included in social media companies' routine transparency reports (quote from Reset Australia [84]):

> *Annual transparency reports are necessary, that include clear prescriptions around information that must be shared to create meaningful transparency. It also helps create the conditions to compare services and track changes over time in meaningful ways.*

Recommended components included details about (1) the *detection* of content deemed to violate social media companies' community guidelines (eg, number of posts within a specified period, number and proportion detected through automated means, and number and proportion reported by users); (2) the accuracy and error rates for automated review processes; (3) the *types* of content reported and moderated on social media platforms, including specific types of violations of community guidelines (eg, self-harm and child sexual abuse material); (4) the *responses* of social media companies when content is deemed to violate their community guidelines and relevant regulations (eg, time taken for posts to be removed); (5) the number of *take-down orders* issued by regulators and social media companies' responses to these orders (eg, action taken by the company and response times); and (6) the number of out-of-court settlements made.

Theme 3: Collaboration

The third theme, *collaboration*, was documented in 71% (50/70) of the reports and included recommendations that social media companies collaborate with important stakeholders at all stages of service design and delivery. This included social media companies consulting and collaborating with governments, relevant agencies, young people, health practitioners, and researchers.

In some reports (12/50, 24%), the authors advocated for investment in research regarding possible impacts of social media on young people's mental health, well-being, and brain development [52,91]. The need for research funding was highlighted in 24% (12/50) of the reports [92-94], along with acknowledgment that conducting such research requires cross-sector cooperation and facilitating researcher access to social media user data. Indeed, several reports (37/50, 74%) highlighted that enhanced collaboration (ie, among social media companies, researchers, and policy makers) will require wider data sharing agreements [33,65,77] (quote from ReachOut, Beyond Blue, and Black Dog Institute [52]):

> *Ensuring that young people's experiences, agency and voices are central to policy development will lead to more relevant and effective interventions.*

Several recommendations emphasized the need to better understand the digital experiences and needs of specific vulnerable groups, who are often at risk of digital harms [86]; hear from a diverse range of voices; and prioritize cross-cultural collaboration [95,96]. For example, this includes engaging with young boys who encounter misogynistic content in digital spaces [95] in addition to engaging with young people from ethnic

minority groups, LGBTQ+ young people, and young people experiencing mental ill-health [97]. As outlined in the following quote, marginalized young people may be particularly vulnerable to digital harms; however, they may also find highly affirming networks on these digital spaces that *support* their well-being (quote from Madden et al [97]):

> *Youth, especially youth of color, LGBTQ+ youth, and young people with depressive symptoms, are in need of safe spaces, support, and resources both online and offline. Young people who turn to social media for support and search for resources online suggest that their offline communities and environments do not provide enough for their needs...*

Some reports (7/50, 14%) highlighted the importance of communication and consistency *between* different social media platforms given that the vast majority of users use multiple platforms [41,78]. This includes cross-sector research and collaboration (eg, developing protocols for age restriction; see theme 5), which should be subject to independent evaluation and auditing (quote from O'Connor [41]):

> *Partnerships between companies can help to recognise and react to known terrorist threats more quickly than otherwise possible and reflect the cross-platform nature of online threats. However, any such effort should be transparent and clearly defined, making the objectives, processes and outcomes of such collaboration clear from the outset, with opportunities for auditing and evaluation by independent experts from civil society and the research community.*

Theme 4: Safety by Design

Overview

The theme safety by design outlines recommendations on how social media companies can design, develop, and deploy products that are safe for young users of their services and platforms. This includes the collection and use of user data and ensuring that potential harms to users are minimized, identified, and addressed. This theme contained 3 subthemes: *privacy*, *content moderation*, and *autonomy*, each of which represents key areas of responsibility for social media companies to deliver safe products to end users.

Theme 4a: Privacy

In total, 21% (15/70) of the reports highlighted the importance of social media companies having suitable processes, policies, and actions regarding the collection and use of user data. Several reports (11/15, 73%) strongly recommended enhanced and updated privacy (eg, establishing the user privacy settings to the maximum by default) and data protection legislation to limit the collection and use of young people's data. This included limiting social media companies in using user data to shape personalized algorithms and advertising to young people [33,77,98,99]. Such legislation could support data minimization (collecting the minimum amount of personal user data to deliver services), reduce personalized content, and create special protections for particularly sensitive data (eg, biometric

information and precise location data), particularly for young users [98,99].

Theme 4b: Content Moderation

A total of 26% (18/70) of the reports explored recommendations for content moderation policies, practices, and design features used by social media companies. Recommendations advocated for stronger investment in content moderation staffing and technologies [78], increased use of human moderation [33,100], and use of advanced AI for detecting and removing content that violates community guidelines [86]. The moderation of content posted across an array of languages was also recommended [84,86] (quote from Amnesty International [33]):

> *Ensure consistency in content moderation decision making, ensure adequate human oversight of automated content moderation and appropriate investment in content moderation resourcing across all languages.*

It was also recommended that social media companies do more to understand how users evade detection in the current systems (eg, using visual cues such as emojis, replacing blacklisted terms with similar-sounding words, or altering posts from their original forms [41,101]) (quote from ISD [41]):

> *TikTok must regularly update its list of blocked terms, incorporating a significantly increased number of variations and misspellings, and apply these filters not just to search results but also to hashtag suggestions.*

Theme 4c: Autonomy

A number of reports (20/70, 29%) provided recommendations pertaining to social media companies increasing the capacity for platform users to make informed decisions regarding their experiences on the platform. Recommendations advocated for safety design features (eg, opting out [102]), allowing young people to be active agents in personalizing their feed or algorithm, including their increased capacity to "suppress" types of content that they do not wish to see on their feeds [33,56,57,89,103]. This design feature may allow users to restrict or limit posts that negatively impact their mental health (eg, disordered eating, misogynistic content, and self-harm) on their feeds [85,94]. Offering young people greater choice about the type of content shown on their feeds is considered a potential protective factor against mental ill-health [89,100] (quote from The US Surgeon General's Advisory [94]):

> *Give users [young people] opportunities to control their online activity by opting out of the content they may find harmful...such as ads involving violence, alcohol, or gambling, or content related to eating disorders.*

Recommendations regarding specific design features on social media platforms were also offered, for example, allowing users greater choice and control over design features (eg, turning off or disabling the "infinite scroll" feature on TikTok [52,98,100]) and greater choice on how their collected data are used (eg, providing informed consent [77,85]). This enhanced user choice may allow young people to use social media platforms in ways

that better meet their needs and preferences [98]. Others (10/20, 50%) offered stronger recommendations. For example, the Office of the Minnesota Attorney General Keith Ellison [102] proposed that "dark patterns" within platform design that encourage excessive use for young people "beyond their explicit desires"—such as infinite scrolling, auto-play videos, engagement-driven algorithms, and frequent notifications—be banned altogether.

Theme 5: Restricting Young People's Access to Social Media

Several reports (14/70, 20%) discussed the potential benefits and drawbacks of implementing age restrictions on young people so that they cannot access social media platforms until they reach specified ages. Some authors (5/14, 36%) recommended the introduction of mandatory age verification for all social media accounts to ensure that users met the minimum age criteria [82,104,105]. However, others (5/14, 36%) highlighted the potential for age verification technologies to exacerbate existing concerns, such as social media companies' collection and use of large amounts of user data [33,98,99]. For example, introducing age verification and limiting young people's access to social media may pose further risks to young people [98]. In particular, age verification will likely involve social media companies assuming responsibility for collection of additional personal user information by requiring government-approved IDs (or other similar documentation) to verify user age, therefore posing privacy concerns (quote from Forland et al [98]):

> While more e orts are needed to ensure children can safely and securely access online spaces, age verification mandates may actually pose more risks than benefits—resulting in unintended consequences for the constitutional rights, privacy, and security of all users.

Thus, recommendations emphasized the necessity for multipronged solutions given that "age verification is no substitute for privacy protections and increased user transparency and control" [98]. Reports also called for evaluation of the effectiveness of age verification technologies as these necessarily mediate any widespread age-based ban [98,100]. In addition, recommendations for alternative approaches to age verification via government IDs were suggested to circumvent privacy concerns [106]. For example, the Australian government suggested taking advantage of an existing service, the Australian Digital ID program, to facilitate age verification in ways that minimize privacy concerns [107]. Other recommendations included involving caregivers in verifying children's ages [104,105], using facial recognition technologies (eg, Yoti [106]), and collaboration among a range of stakeholders to develop best practices and effective age verification systems to minimize the impact on young users [94,98,107] (quote from Forland et al [98]):

> Insights from industry, civil society, regulators, and users of all ages should be taken into consideration to create standardized best practices and protocols for age verification.

Finally, recommendations were made to explore alternatives to age verification processes (eg, stronger legislation, empowered regulators, and safety- and security-by-design features [98,104,106,107]) to mitigate the additional privacy and safety concerns (quote from Johnson [106]):

> An alternative to age verification would require device operating systems to create a "trustworthy child flag" for user accounts that signals to apps and websites that a user is underage and require apps and websites that serve age-restricted content to check for this signal from their users and block underage users from this content.

Discussion

Principal Findings

Overview

To our knowledge, this is the first review to synthesize recommendations regarding social media and young people's mental health. This synthesis is especially timely in the context of ongoing discourse surrounding harms associated with social media, legislative changes regarding young people's access to social media [70,136], and growing tension between governments and social media companies [137]. This review included 70 reports from Australia, Canada, the United Kingdom, and the United States. The identified themes highlight wide-ranging opportunities for governments, regulators, and stakeholder collaboration in ensuring that social media products are designed and delivered to young people in ways that safeguard their mental health. Moreover, these themes are consistent with published empirical works and reviews [7,8,23,36].

Strategies for Limiting Harmful Content on Social Media Platforms

While social media companies have introduced various content moderation tools to reduce harmful content on their platforms [138], many reports (10/18, 56%) highlighted that existing strategies are insufficient to prevent young people from being exposed to digital harms and advocated strongly for improved content moderation practices and policies [30,33,104]. This aligns with empirical evidence that indicates that more than 3 in 4 young people report exposure to digital hate content [4,23] and findings indicating that young people are approximately 4 times more likely to see posts related to self-harm or suicide relative to those aged >25 years [11]. Perspectives on moderation strategies were mixed, with some (6/18, 33%) advocating for increased investment in human moderators [33,100], whereas others (5/18, 28%) recommended greater use of AI-facilitated moderation. This comes in the wake of many social media companies' retrenchment of content moderation workforces in recent years (eg, Meta's decision to replace paid moderators with "voluntary community notes" on its platforms [139-142]). However, as noted throughout the literature, overreliance on AI-facilitated moderation and natural language processing techniques warrants careful consideration [44]. AI moderation tools often fail to recognize and remove iterations of banned material (eg, terms such as "unaliving" or hashtags such as #suwerslide instead of #suicide) [88]. Indeed, users routinely circumventing banned words and topics only serves

to highlight important gaps in the efficacy of current moderation practices [46]. Notably, the included reports rarely alluded to young people's *intentional* exposure to potentially harmful digital material, such as posts related to self-harm or suicide. This is a considerable oversight as previous studies have shown that some young people actively seek information and support regarding these digital challenges, particularly given that social media can offer anonymity, cost-effectiveness, and a sense of connection with others experiencing similar hardships [11]. Nonetheless, the need for effective and sophisticated moderation tools to reduce exposure to potentially harmful posts was emphasized across the reports, as has been advocated for widely in the extant academic literature [44,46]. Such action is likely to garner support and foster trust in social media services among the public based on data indicating that the general community is supportive of governments and social media companies taking proactive steps to remove harmful digital content [23]. However, there is also a need to consider strategies for making posts related to mental health available and safe for young people to access on the web [11].

Recommendations also emphasized that social media companies must be held responsible for young users' exposure to harmful digital content through legislative enforcement of compliance. Strong support was expressed for legislative changes to support ensuring the liability of social media companies for harmful content being created and disseminated on their platforms rather than allowing for immunity [77,78]. Alongside amending and introducing necessary legislation and expanding the definition of online harms [80,108,109], the authors recommended empowering independent regulators with meaningful auditing, investigative, and enforcement powers [65,77].

Crucially, regulations and moderation practices must be fit for purpose and informed by a thorough understanding of digital technologies (including routinely used strategies for nondetection). They should be continually reviewed and adapted to ensure that they are responsive to technological advancements and emerging digital safety issues, such as the use of "bot accounts," AI-generated content, and encoded file formats, and responsive to changing digital landscapes to avoid the "legal lag" [65].

Strategies for Improving Social Media Platform Design

This review captured several recommendations to improve design features that keep young people engaged on social media platforms for long periods, such as infinite scrolling, auto-play videos upon opening platforms, and constant push notifications [56,100]. Given young people's developmental stage, these features are considered particularly harmful to them [143] and have been associated with feelings of depression, anxiety, hopelessness, and isolation [144,145]. As a result, the reports recommended that social media platforms use human-centered design approaches, where users (including young people) and other stakeholders (eg, caregivers) are involved in platform design [100].

Implementing age verification in social media platforms has been a topic of considerable debate, with Australia recently mandating age restrictions for platforms such as Instagram, TikTok, and Snapchat [10,70]. While some (5/14, 36%) argued

that limiting young people's access to social media is crucial for preventing their exposure to inappropriate content [3,146], others (5/14, 36%) expressed concerns about feasibility and the possible risks that age verification may present [98]. Criticism of age verification included uncertainty about how to effectively implement verification measures, risk of underage users bypassing restrictions, and privacy concerns. For example, if age verification measures require users to provide identification documents to sign up to social media platforms, they may risk further infringements on users' personal information or serious privacy breaches [147]. Consequently, decisions regarding age restriction and verification require careful consideration regarding how best to ensure user safety, accessibility, and privacy [100].

Age restriction represents one of many possible responses to the perceived risks to young people's mental health that social media can pose. However, this approach ignores the possible benefits and adaptive uses of social media among young people [7-9], including social connection and self-expression [6]. Indeed, based on existing literature, age restrictions could have significant *negative* impacts on some young people, particularly marginalized groups such as LGBTQ+ and migrant youth [146]. Many of these young people rely on social media to connect with their communities, obtain information they would not seek (or be able to seek) in offline contexts, combat stigma and isolation, and find support [9,11]. Some have also noted that removing young people's access to social media risks them turning to digital spaces that have less stringent safety measures and pose *greater* risk to mental ill-health, such as unmoderated forums [148]. As such, restrictive approaches have been described by experts and academics as a "blunt tool" [147,149] that does little to ensure that social media products are made safer for all users, including young people. In sum, while some view age restriction as a necessary step toward safeguarding young people, others advocate against this and propose more nuanced approaches—such as implementing and legislating age-appropriate safety-by-design features and content moderation practices. The latter approach may offer a greater balance between safeguarding youth and preserving their access to supportive digital spaces [110,147].

Strategies for Increasing Transparency and Collaboration

The need for greater transparency from social media companies was frequently highlighted in the recommendations. This included communicating openly and clearly with governments, regulators, users, and the public about their policies and practices regarding the type of data that are collected from users and how they are used, how algorithms are shaped, what posts constitute paid or promotional content, and the types and frequencies of posts being removed from these digital spaces [110].

Although several studies (10/16, 62%) suggested that social media companies release regular transparency reports [41,87,90], this recommendation has since been implemented in countries such as Australia, the United States, the United Kingdom, and Canada. To combat regulatory scrutiny, social media platforms release quarterly transparency reports that highlight the measures taken to address misinformation and disinformation [84].

However, regulators have raised concerns over the quality and utility of current transparency reports [86]. For example, when figures about violative content are presented (eg, 10,000 posts were removed), these figures are often a combination of many metrics and lack clarity about the *type* of violation for which the posts were removed (eg, posts inciting violence or posts proven to be false after fact-checking [84]). They also do not indicate how many reported posts are not removed, the maximum time until removal, and the scope of the posts' spread before their removal. Additional information is required to increase the utility of transparency reports, such as the number of take-down orders issued by regulators, median and maximum times until posts deemed to violate community guidelines are removed, and specific actions taken following removal [84,150].

Transparency reports have also been criticized for their lack of objectivity as they are often developed by the social media companies that have a vested interest in the findings. In addition, the social media companies' refusal to grant data access to researchers prevents independent verification of the claims presented in these reports. Indeed, authors of the included reports argued that providing approved researchers with access to the data collected by social media platforms will enable the evaluation and auditing of content moderation systems, platform policies, and their alignment with policies and regulations [84,87,98]. This echoes similar calls by academics [151]. Researcher access may also enable the auditing of actions taken by social media companies against harmful content and users who violate platform policies [92]. To minimize privacy risks and maximize user safety, access to public data should only be made available to "vetted" researchers [84,151] and will require data sharing agreements in addition to coordination and collaboration between social media companies and relevant regulators [100].

Currently, there is a significant gap in longitudinal research on the effects of social media use on young people, leading to an incomplete understanding of both potential harms and benefits [3]. Indeed, some studies suggest that social media use is *positively* associated with important correlates of well-being such as loneliness [13]. As highlighted in previous reviews, long-term studies are essential to comprehensively assess how particular social media content and behaviors impact youth mental health, development, and behavior over time [7,9]. Academics also increasingly emphasize the importance of considering nuances in *how* young people use social media (eg, which platforms and the types of content viewed) and for what purposes (eg, entertainment, information gathering, and social connection) rather than framing proposed links between social media use and mental health as a linear relationship, wherein increased time on social media necessarily predicts poorer outcomes [6,13]. Furthermore, there is a need to consider whether particular groups are at elevated risk of being impacted by potential digital harms (eg, mental ill-health, exploitation, and disinformation) as this may guide targeted prevention and intervention efforts. Providing researchers with access to real-time data, alongside substantial investment in research, would greatly enhance the ability to conduct such studies, enabling the monitoring of changes and trends in young people's social media behavior and broader effects over extended periods

[92]. Importantly, researchers also recommend that empirical findings be coupled with young people's insights and perspectives, which can be obtained via qualitative studies and meaningful consultation [6,11]. These approaches would offer more informed conclusions and evidence-based recommendations for managing the impacts of social media on young people [87].

While collaboration among regulators, social media companies, and researchers is a crucial step toward creating safer platforms, young people should remain at the center of this collaboration. Several reports (10/50, 20%) recommended consulting and collaborating with young people, including those who are typically underrepresented during the development of legislation, regulation, and social media platform design [33,39,111]. Existing research shows that policies and programs that involve young people in their development are more likely to be effective and well received [52]. This participatory approach also ensures that the solutions are more relevant and tailored to the specific needs and experiences of young people [52]. Young users are often the first to encounter new and evolving risks on social media platforms. Therefore, their input is invaluable in identifying and addressing these challenges. By incorporating their perspectives, policies can be better equipped to tackle the complexities of digital behavior, mental health concerns, and digital safety, offering more nuanced and practical solutions that align with young people's lived experiences [52,111].

Limitations

This review's findings should be viewed in light of several limitations, including the paucity of non–English-language documents. This limits the generalizability of the results to non–English-speaking contexts and may not account for the impact of social media practice and policy on youth mental health worldwide. While this review focused on countries that have been leading in terms of regulatory efforts and public discourse regarding social media and youth mental health, it is suggested that future international reviews broaden their scope and capture recommendations for governments and regulators and the social media industry at a global level. A global review will also bring forward the issues faced by middle- to low-income countries where young people are more likely to be vulnerable to the negative impacts of social media (eg, lack of resources to circumvent targeted advertising and marketing). In addition, although this is the first review to collate recommendations regarding social media and youth mental health, the recommendations were not categorized based on the developmental stages of users aged 12 to 25 years. This omission was due to the absence of age-specific recommendations in the existing literature. Future research should be mindful of sensitizing recommendations based on user age as blanket policies that attempt to address the unique needs of both those aged 12 years and those aged 25 years under the same regulatory framework may prove ineffective.

This review predominantly captured gray literature, which precluded evaluation of the quality of the reports and contributed to the widespread lack of evidence-based recommendations. Nonetheless, the recommendations were rationally derived by subject matter experts and organizations. Our focus on the

available (largely gray) literature represents a pragmatic attempt to synthesize recommendations that are likely to be viewed by governments and regulators in the absence of corresponding and contemporaneous academic literature. Furthermore, given our focus on recommendations for the social media industry and regulators, it was beyond the scope of this review to collate recommendations specifically for young people, clinicians, caregivers, or educators. While we acknowledge that this may have meant that we overlooked some important recommendations for protecting young people's mental health and digital safety, our focus was intentional to highlight the many and varied opportunities for the social media industry itself, and associated regulators, to enforce safety. While the authors acknowledge that this overlooks some relevant recommendations regarding the relationships between young people's mental health and their social media use, the goal of this review was to move away from user responsibility and demonstrate the varied opportunities for the social media industry and regulators.

Finally, the recommendations in this review were rarely informed by rigorous peer-reviewed evidence, which requires amelioration as the evidence base proliferates. This limits current understanding regarding the efficacy and impacts of enacting the types of recommendations proposed in the included reports. Rigorous, peer-reviewed studies are needed to systematically evaluate the efficacy and real-world impacts of the proposed recommendations. Strengthening this evidence base will ultimately guide more informed decision-making and policy development.

Conclusions

This scoping review aimed to provide a synthesis of published recommendations for governments, regulators, and the social media industry at a time when both government bodies and social media companies are considering implementing landmark changes to protect the mental health of young people. The scope of evidence across the 5 interrelated themes highlighted in this review emphasizes the range of options available to both industry and government and the imperative to undertake a multipronged approach. These 5 themes, while distinct, are clearly interrelated and collectively offer a comprehensive array of strategies to address the increasing presence of social media in young people's lives and likely their health. At a time when the empirical evidence for the harms and benefits of social media in young people is still emerging, the expert-based recommendations in this review establish a tangible path forward for both governments and social media companies to safeguard the health of young people. This review underscores the pressing need for coordinated, evidence-informed action from policy makers and platforms, particularly as discourse surrounding social media and youth mental health continues to intensify and outpace regulatory and industry responses. In this context, the emergence of reactive and restrictive action, such as proposals to ban young people from social media, risks diverting attention from the more complex but essential task of creating safer and more supportive digital environments.

Authors' Contributions

JC was responsible for project administration. JC and VP led the project conceptualization, data curation (including developing the search strategy and inclusion and exclusion criteria), investigation, formal analysis, writing—original draft, and writing—review and editing. RB and MJW contributed to writing—review and editing and visualization. LLS contributed to writing—review and editing. ZS contributed to supervision and writing—review and editing.

Conflicts of Interest

Multimedia Appendix 1

Characteristics of the included studies.
[DOCX File , 66 KB-Multimedia Appendix 1]

Multimedia Appendix 2

PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Review) checklist.
[PDF File (Adobe PDF File), 498 KB-Multimedia Appendix 2]

References

1. Young people's social media use-what impact does it have? ANU Centre for Social Research & Methods. URL: https://generationsurvey.org.au/data_story/social-media-use-and-life-satisfaction/ [accessed 2025-05-29]
2. Children and parents: media use and attitudes report. Ofcom. URL: https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/media-literacy-research/children/children-media-use-and-attitudes-2024/childrens-media-literacy-report-2024.pdf?v=368229 [accessed 2025-01-31]
3. Social media and youth mental health. The U.S. Surgeon General's Advisory. URL: https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf [accessed 2025-01-31]
4. Young people and exposure to harmful online content in 2022. Statistics Canada. URL: https://www150.statcan.gc.ca/n1/pub/11-627-m/11-627-m2024005-eng.htm [accessed 2025-01-31]

5. McGorry PD, Mei C, Dalal N, Alvarez-Jimenez M, Blakemore S, Browne V, et al. The Lancet psychiatry commission on youth mental health. Lancet Psychiatry. Sep 2024;11(9):731-774. [doi: 10.1016/S2215-0366(24)00163-9] [Medline: 39147461]

6. Hamilton JL, Nesi J, Choukas-Bradley S. Reexamining social media and socioemotional well-being among adolescents through the lens of the COVID-19 pandemic: a theoretical review and directions for future research. Perspect Psychol Sci. May 10, 2022;17(3):662-679. [FREE Full text] [doi: 10.1177/17456916211014189] [Medline: 34756118]

7. Khalaf AM, Alubied AA, Khalaf AM, Rifaey AA. The impact of social media on the mental health of adolescents and young adults: a systematic review. Cureus. Aug 2023;15(8):e42990. [FREE Full text] [doi: 10.7759/cureus.42990] [Medline: 37671234]

8. Best P, Manktelow R, Taylor B. Online communication, social media and adolescent wellbeing: a systematic narrative review. Child Youth Serv Rev. Jun 2014;41:27-36. [doi: 10.1016/j.childyouth.2014.03.001]

9. Berger MN, Taba M, Marino JL, Lim MS, Skinner SR. Social media use and health and well-being of lesbian, gay, bisexual, transgender, and queer youth: systematic review. J Med Internet Res. Sep 21, 2022;24(9):e38449. [FREE Full text] [doi: 10.2196/38449] [Medline: 36129741]

10. Tipping the balance. LGBTQI+ teens' experiences negotiating connection, self-expression and harm online. eSafety Commissioner. URL: https://www.esafety.gov.au/sites/default/files/2024-06/Tipping-the-balance-June-2024.pdf [accessed 2025-05-29]

11. Robinson J, La SL, Kenny B, Cooper C, Lamblin M, Spittal M. How do Australian social media users experience self-harm and suicide-related content? A national cross-sectional survey comparing young people and adults. PsyArXiv. Preprint published online July 29, 2024. [FREE Full text] [doi: 10.31234/osf.io/5rh93]

12. Dictionary of psychology - mental health. American Psychological Association. 2018. URL: https://dictionary.apa.org/mental-health [accessed 2025-01-31]

13. Dissing AS, Hulvej Rod N, Gerds TA, Lund R. Smartphone interactions and mental well-being in young adults: a longitudinal study based on objective high-resolution smartphone data. Scand J Public Health. May 2021;49(3):325-332. [doi: 10.1177/1403494820920418] [Medline: 32536319]

14. Dyer WJ, Coyne SM, Gale M, Sheppard JA. Who's most at risk? A person-centered approach to understanding the long-term relationship between early social media use and later depression across adolescence. J Adolesc. Oct 26, 2024;96(7):1555-1568. [doi: 10.1002/jad.12362] [Medline: 38924113]

15. García-Manglano J, Fernández A, Serrano C, López-Madrigal C, Fernández-Duval G, de la Rosa Fernández-Pacheco P, et al. Social media and mental health: the role of interpersonal relationships and social media use motivations, in a nationally representative, longitudinal sample of Spanish emerging adults. J Soc Pers Relat. Mar 26, 2024;41(5):1157-1182. [doi: 10.1177/02654075241230248]

16. Hall J, Pennington N, Merolla A. Which mediated social interactions satisfy the need to belong? J Comp Med Comm. 2022;28(1):zmac026. [doi: 10.1093/jcmc/zmac026]

17. Wenninger H, Krasnova H, Buxmann P. Understanding the role of social networking sites in the subjective well-being of users: a diary study. Eur J Inf Syst. Aug 02, 2018;28(2):126-148. [doi: 10.1080/0960085X.2018.1496883]

18. Vilhelmsson A, Svensson T, Meeuwisse A. Mental ill health, public health and medicalization. Public Health Ethics. Oct 28, 2011;4(3):207-217. [doi: 10.1093/phe/phr030]

19. Kruzan K, Nesi J, Hamilton J. Media influences on self-harm, suicidality, and suicide. In: Christakis DA, Hale L, editors. Handbook of Children and Screens: Digital Media, Development, and Well-Being from Birth Through Adolescence. Cham, Switzerland. Springer; 2025:141-147.

20. Hall JA. Ten myths about the effect of social media use on well-being. J Med Internet Res. Nov 25, 2024;26:e59585. [FREE Full text] [doi: 10.2196/59585] [Medline: 39586087]

21. What parents worry about: carer concerns for youth mental health and wellbeing. ReachOut Australia. URL: https://d1robvhmkdqpun.cloudfront.net/20ec30143a2719d6d3fda2a71980ea63.pdf [accessed 2025-01-31]

22. What are you worried about?: young people's stress burden and its impacts on their wellbeing. ReachOut Australia. URL: https://d1robvhmkdqpun.cloudfront.net/6ee09aefcd82ab783758e1da8dac6dce.pdf [accessed 2025-01-31]

23. Enoch F, Johansson P, Bright J, Margetts HZ. Tracking experiences of online harms and attitudes towards online safety interventions: findings from a large-scale, nationally representative survey of the British public. SSRN Journal. preprint posted online April 24, 2023. [FREE Full text] [doi: 10.2139/ssrn.4416355]

24. Beyari H. The relationship between social media and the increase in mental health problems. Int J Environ Res Public Health. Jan 29, 2023;20(3):2383. [FREE Full text] [doi: 10.3390/ijerph20032383] [Medline: 36767749]

25. Dyer C. Social media content contributed to teenager's death "in more than a minimal way," says coroner. BMJ. Oct 03, 2022;379:o2374. [doi: 10.1136/bmj.o2374] [Medline: 36191932]

26. Choukas-Bradley S, Roberts SR, Maheux AJ, Nesi J. The perfect storm: a developmental-sociocultural framework for the role of social media in adolescent girls' body image concerns and mental health. Clin Child Fam Psychol Rev. Dec 16, 2022;25(4):681-701. [FREE Full text] [doi: 10.1007/s10567-022-00404-5] [Medline: 35841501]

27. Wells G, Horwitz J, Seetharaman D. Facebook knows Instagram is toxic for teen girls, company documents show. The Wall Street Journal. URL: https://tinyurl.com/mr2df5z2 [accessed 2025-01-31]

28. Thai H, Davis CG, Mahboob W, Perry S, Adams A, Goldfield GS. Reducing social media use improves appearance and weight esteem in youth with emotional distress. Psychol Pop Media. Jan 2024;13(1):162-169. [doi: 10.1037/ppm0000460]

29. Griffiths S, Harris EA, Whitehead G, Angelopoulos F, Stone B, Grey W, et al. Does TikTok contribute to eating disorders? A comparison of the TikTok algorithms belonging to individuals with eating disorders versus healthy controls. Body Image. Dec 2024;51:101807. [FREE Full text] [doi: 10.1016/j.bodyim.2024.101807] [Medline: 39504757]

30. Lonergan AR, Bussey K, Fardouly J, Griffiths S, Murray SB, Hay P, et al. Protect me from my selfie: examining the association between photo-based social media behaviors and self-reported eating disorders in adolescence. Int J Eat Disord. May 07, 2020;53(5):485-496. [doi: 10.1002/eat.23256] [Medline: 32259344]

31. Body image: we are more than what we look like. Scottish Government's Body Image Advisory Group on Good Body Image & Mental Health Foundation. URL: https://www.mentalhealth.org.uk/sites/default/files/2022-06/MHF-Scotland-Body-image-report.pdf [accessed 2025-01-31]

32. Digital lives of Aussie teens. eSafety Commissioner. URL: https://www.esafety.gov.au/research/digital-lives-of-aussie-teens [accessed 2025-01-31]

33. Driven into darkness: how TikTok's 'for you' feed encourages self-harm and suicidal ideation. Amnesty International. URL: https://www.amnesty.org/en/documents/POL40/7350/2023/en/ [accessed 2025-01-31]

34. Finkelhor D, Turner H, Colburn D. Prevalence of online sexual offenses against children in the US. JAMA Netw Open. Oct 03, 2022;5(10):e2234471. [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.34471] [Medline: 36239942]

35. Finkelhor D, Turner H, Colburn D. Which dynamics make online child sexual abuse and cyberstalking more emotionally impactful: perpetrator identity and images? Child Abuse Negl. Mar 2023;137:106020. [doi: 10.1016/j.chiabu.2023.106020] [Medline: 36657227]

36. Paat Y, Markham C. Digital crime, trauma, and abuse: internet safety and cyber risks for adolescents and emerging adults in the 21st century. Soc Ment Health. Nov 13, 2020;19(1):18-40. [doi: 10.1080/15332985.2020.1845281]

37. Gewirtz-Meydan A, Mitchell K, O?Brien J. Sexual posttraumatic stress among investigators of child sexual abuse material. J Policy Pract. 2023;17:paad052. [doi: 10.1093/police/paad059]

38. Infodemics and misinformation negatively affect people's health behaviours: new WHO review finds. World Health Organization. URL: https://tinyurl.com/556fjhyy [accessed 2025-01-31]

39. Dorn M, Bundtzen S, Schwieter C, Gandhi M. Emerging platforms and technologies: an overview of the current threat landscape and its policy implications. Institute for Strategic Dialogue. URL: https://www.isdglobal.org/wp-content/uploads/2023/10/Emerging-Platforms-and-Technologies-An-Overview-of-the-Current-Threat-Landscape-and-its-Policy-Implications.pdf [accessed 2025-01-31]

40. Borges do Nascimento IJ, Beatriz Pizarro A, Almeida J, Azzopardi-Muscat N, André Gonçalves M, Björklund M, et al. Infodemics and health misinformation: a systematic review of reviews. Bull World Health Organ. Sep 01, 2022;100(9):544-561. [doi: 10.2471/blt.21.287654]

41. O'Connor C. Hatescape: an in-depth analysis of extremism and hate speech on TikTok. Institute for Strategic Dialogue. URL: https://www.isdglobal.org/wp-content/uploads/2021/08/HateScape_v5.pdf [accessed 2025-05-29]

42. Johnson NF, Leahy R, Restrepo NJ, Velasquez N, Zheng M, Manrique P, et al. Hidden resilience and adaptive dynamics of the global online hate ecology. Nature. Sep 21, 2019;573(7773):261-265. [doi: 10.1038/s41586-019-1494-7] [Medline: 31435010]

43. O'Donnell C, Shor E. "This is a political movement, friend": why "incels" support violence. Br J Sociol. Mar 16, 2022;73(2):336-351. [doi: 10.1111/1468-4446.12923] [Medline: 35034349]

44. Gongane VU, Munot MV, Anuse AD. Detection and moderation of detrimental content on social media platforms: current status and future directions. Soc Netw Anal Min. Sep 05, 2022;12(1):129. [FREE Full text] [doi: 10.1007/s13278-022-00951-3] [Medline: 36090695]

45. Engstrom E, Feamster N. The limits of filtering: a look at the functionality and shortcomings of content detection tools. Engine. URL: https://www.engine.is/the-limits-of-filtering [accessed 2025-01-31]

46. Gerrard Y. Beyond the hashtag: circumventing content moderation on social media. New Media Soc. May 28, 2018;20(12):4492-4511. [doi: 10.1177/1461444818776611]

47. Stewart E. Detecting fake news: two problems for content moderation. Philos Technol. Feb 11, 2021;34(4):923-940. [FREE Full text] [doi: 10.1007/s13347-021-00442-x] [Medline: 33589871]

48. Position paper - recommender systems and algorithms. eSafety Commissioner. URL: https://www.esafety.gov.au/sites/default/files/2022-12/Position%20statement%20-%20Recommender%20systems%20and%20algorithms.pdf [accessed 2025-01-31]

49. Metzler H, Garcia D. Social drivers and algorithmic mechanisms on digital media. Perspect Psychol Sci. Sep 19, 2024;19(5):735-748. [FREE Full text] [doi: 10.1177/17456916231185057] [Medline: 37466493]

50. What is infinite scrolling? Interaction Design Foundation. URL: https://www.interaction-design.org/literature/topics/infinite-scrolling [accessed 2025-01-31]

51. Montag C, Lachmann B, Herrlich M, Zweig K. Addictive features of social media/messenger platforms and freemium games against the background of psychological and economic theories. Int J Environ Res Public Health. Jul 23, 2019;16(14):2612. [FREE Full text] [doi: 10.3390/ijerph16142612] [Medline: 31340426]

52. Inquiry into social media impacts on Australian Society. Joint Select Committee on Social Media and Australian Society Submission 168. URL: https://www.blackdoginstitute.org.au/wp-content/uploads/2024/07/Sub168_ReachOut-Beyonce-Blue-BDI.pdf [accessed 2025-01-31]

53. Yang C, Xu X, Nunes BP, Siqueira SW. Bubbles bursting: investigating and measuring the personalisation of social media searches. Telemat Inform. Aug 2023;82:101999. [doi: 10.1016/j.tele.2023.101999]

54. Farthing R. Designing for disorder: algorithms amplify pro-anorexia content to teens and children as young. Reset Australia. URL: https://au.reset.tech/uploads/insta-pro-eating-disorder-bubble-april-22-1.pdf [accessed 2025-01-31]

55. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science. Mar 09, 2018;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]

56. Adolescent screen use and mental health: summary of findings from the Future Proofing Study. Black Dog Institute. URL: https://www.blackdoginstitute.org.au/research-projects/teens-screens-adolescent-mental-health-screen-use/ [accessed 2025-05-29]

57. Teenagers and social media. ReachOut Australia. URL: https://parents.au.reachout.com/staying-safe-online/social-media/teenagers-and-social-media [accessed 2025-01-31]

58. Thomas E, Balint K. Algorithms as a weapon against women: how YouTube lures boys and young men into the "manosphere". Institute for Strategic Dialogue & Reset Tech. URL: https://www.isdglobal.org/wp-content/uploads/2022/04/Algorithms-as-a-weapon-against-women-ISD-RESET.pdf [accessed 2025-01-31]

59. Online safety act 2021. Australian Government. URL: https://www.legislation.gov.au/Details/C2021A00076 [accessed 2025-01-31]

60. A guide to the online safety bill. Government UK. URL: https://www.gov.uk/guidance/a-guide-to-the-online-safety-bill [accessed 2025-01-31]

61. Children's online privacy protection rule ("COPPA"). US Government. 1998. URL: https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa [accessed 2025-01-31]

62. S.3663 - kids online safety act. United States Government. URL: https://www.congress.gov/bill/117th-congress/senate-bill/3663/text [accessed 2025-05-29]

63. Implementing the online news act. Government of Canada. URL: https://crtc.gc.ca/eng/industr/info.htm [accessed 2025-01-31]

64. 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. Front Child Adolesc Psychiatry. Dec 15, 2023;2:1274263. [FREE Full text] [doi: 10.3389/frcha.2023.1274263] [Medline: 39839592]

65. A duty of care in Australia's online safety act. Reset Australia. URL: https://au.reset.tech/uploads/Duty-of-Care-Report-Reset.Tech.pdf [accessed 2025-01-31]

66. La Sala L, Sabo A, Michail M, Thorn P, Lamblin M, Browne V, et al. Online safety when considering self-harm and suicide-related content: qualitative focus group study with young people, policy makers, and social media industry professionals. J Med Internet Res. Mar 10, 2025;27:e66321. [FREE Full text] [doi: 10.2196/66321] [Medline: 40063940]

67. Children's experiences of legal but harmful content online. National Society for the Prevention of Cruelty to Children. URL: https://learning.nspcc.org.uk/media/pi1hqvez/legal-but-harmful-content-online-helplines-insight-briefing.pdf [accessed 2025-01-31]

68. Vulnerable connections: expert panel on public safety in the digital age. Council of Canadian Academies (CCA). URL: https://www.cca-reports.ca/wp-content/uploads/2023/04/Vulnerable-Connections_FINAL_DIGITAL_EN_UPDATED.pdf [accessed 2025-01-31]

69. Not just algorithms: assuring user safety online with systemic regulatory frameworks. Analysis & Policy Observatory. URL: https://apo.org.au/sites/default/files/resource-files/2024-03/apo-nid326122.pdf [accessed 2025-01-31]

70. Social media reforms to protect our kids online pass Parliament. Prime Minister of Australia. URL: https://www.pm.gov.au/media/social-media-reforms-protect-our-kids-online-pass-parliament [accessed 2025-01-31]

71. Social media ban for under-16s 'on the table' in UK. BBC News. URL: https://www.bbc.com/news/articles/ce9gpdrx829o.amp [accessed 2025-01-31]

72. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. Mar 29, 2021;372:n71. [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]

73. Our mission. Overton. URL: https://www.overton.io/about [accessed 2025-01-31]

74. Gupta K, DSilva MH. Proliferation of social media during the COVID-19 pandemic: a statistical enquiry. J Xi'an Univ Archit Technol. 2020;XII(V):12.

75. Home page. Covidence. URL: https://www.covidence.org/ [accessed 2025-05-29]

76. Turn data complexity into clarity. Lumivero. Jan 29, 2002. URL: https://lumivero.com/ [accessed 2025-05-29]

77. How to make online platforms more transparent and accountable to Canadian users. Canadian Commission on Democratic Expression. URL: https://static1.squarespace.com/static/5ea874746663b45e14a384a4/t/62794492003e16063fab7552/1652114580390/DemX+2+-+English+-+May+4.pdf [accessed 2025-01-31]

78. Investigative report on the role of online platforms in the tragic mass shooting in buffalo on May 14, 2022. Office of the New York State Attorney General Letitia James. URL: https://ag.ny.gov/sites/default/files/buffaloshooting-onlineplatformsreport.pdf [accessed 2025-01-31]

79. Inquiry into social media and online safety submission 37. Eating Disorders Families Australia. URL: https://www.aph.gov.au/Parliamentary_Business/Committees/House/Former_Committees/Social_Media_and_Online_Safety/SocialMediaandSafety/Submissions [accessed 2025-01-31]

80. Patnaik S, Litan RE. TikTok shows why social media companies need more regulation. Center on Regulation & Markets. URL: https://www.brookings.edu/wp-content/uploads/2023/05/20230511_CRM_PatnaikLitan_TikTok_FINAL.pdf [accessed 2025-01-31]

81. Social media, body image, and eating disorders. Butterfly Foundation. URL: https://butterfly.org.au/wp-content/uploads/2024/05/Roundtable-Recommendations-FINAL-ONLINE.pdf [accessed 2025-01-31]

82. Changing the perfect picture: an inquiry into body image. Women and Equalities Committee. URL: https://dera.ioe.ac.uk/id/eprint/37748/1/download%20%281%29.pdf [accessed 2025-01-31]

83. Singh A, Jooshandeh J. Platforms and the public square: a taxonomy of misinformation and the misinformed. Royal Society for Arts & BT Group. URL: https://www.thersa.org/globalassets/_foundation/new-site-blocks-and-images/reports/2021/11/platforms_and_the_public_square.pdf [accessed 2025-01-31]

84. Achieving digital platform public transparency in Australia. Reset Australia. URL: https://au.reset.tech/Digital-platform-public-transparency.pdf [accessed 2025-01-31]

85. 'I feel exposed': caught in TikTok's surveillance web. Amnesty International. URL: https://www.amnesty.org/en/documents/pol40/7349/2023/en/ [accessed 2025-01-31]

86. Bundtzen S. Misogynistic pathways to radicalisation: recommended measures to assess and mitigate online gender-based violence. Institute of Strategic Dialogue. URL: https://www.isdglobal.org/wp-content/uploads/2023/09/Misogynistic-Pathways-to-Radicalisation-Recommended-Measures-for-Platforms-to-Assess-and-Mitigate-Online-Gender-Based-Violence.pdf [accessed 2025-01-31]

87. Online targeting: final report and recommendations. Centre for Data Ethics and Innovation. URL: https://www.gov.uk/government/publications/cdei-review-of-online-targeting/online-targeting-final-report-and-recommendations#introduction [accessed 2025-01-31]

88. Suicide, incels, and drugs: how TikTok's deadly algorithm harms kids. EKO. URL: https://s3.amazonaws.com/s3.sumofus.org/images/eko_Tiktok-Report_FINAL.pdf [accessed 2025-01-31]

89. The Jed Foundation (JED) recommendations for safeguarding youth well-being on social media platforms. The Jed Foundation. URL: https://tinyurl.com/yj3v9rtz [accessed 2025-01-31]

90. Cox K, Ogden T, Jordan V, Paille P. COVID-19, disinformation and hateful extremism. RAND Europe. URL: https://tinyurl.com/jyruzj3j [accessed 2025-01-31]

91. Bradshaw K, Vaillancourt T. Freedom of thought, social media and the teen brain. Centre for International Governance Innovation. URL: https://www.cigionline.org/static/documents/FoT_PB_no.9.pdf [accessed 2025-01-31]

92. Health advisory on social media use in adolescence. American Psychological Association. URL: https://www.apa.org/topics/social-media-internet/health-advisory-adolescent-social-media-use.pdf [accessed 2025-01-31]

93. Dubicka B, Theodosiou L. Technology Use and the Mental Health of Children and Young People. Royal College of Psychiatrists. URL: https://www.rcpsych.ac.uk/docs/default-source/improving-care/better-mh-policy/college-reports/college-report-cr225.pdf [accessed 2025-01-31]

94. Protecting youth mental health. The U.S. Surgeon General's Advisory. URL: https://www.hhs.gov/sites/default/files/surgeon-general-youth-mental-health-advisory.pdf [accessed 2025-01-31]

95. Diepeveen S. How does social media influence gender norms among adolescent boys? Key evidence and policy implications. Advancing Learning and Innovation on Gender Norms (ALiGN). URL: https://www.alignplatform.org/sites/default/files/2024-02/align-socialmedia-briefingnote-jan24-proof05.pdf [accessed 2025-01-31]

96. New York City's role in the national crisis of social media and youth mental health: framework for action. NYC Health. URL: https://www.nyc.gov/assets/doh/downloads/pdf/mh/social-media-youth-mental-health-framework-action.pdf [accessed 2025-01-31]

97. Madden M, Calvin A, Hasse A. A double-edged sword: how diverse communities of young people think about the multifaceted relationship between social media and mental health. Common Sense Media & HopeLab. URL: https://www.commonsensemedia.org/sites/default/files/research/report/2024-double-edged-sword-hopelab-report_final-release-for-web-v2.pdf [accessed 2025-01-31]

98. Forland S, Meysenburg N, Solis E. Age verification: the complicated effort to protect youth online. Open Technology Institute. URL: https://www.newamerica.org/oti/reports/age-verification-the-complicated-effort-to-protect-youth-online/ [accessed 2025-01-31]

99. Farthing R. How outdated approaches to regulation harm children and young people and why Australia urgently needs to pivot. Reset Australia, ChildFund Australia, & Australian Child Rights Taskforce. URL: https://au.reset.tech/uploads/insta-pro-eating-disorder-bubble-april-22-1.pdf [accessed 2025-01-31]

100. Online health and safety for children and youth: best practices for families and guidance for industry. Kids Online Health and Safety Task Force. URL: https://www.samhsa.gov/sites/default/files/online-health-safety-children-youth-report.pdf [accessed 2025-01-31]

101. TikTok series: policy recommendations. Institute for Strategic Dialogue. URL: https://www.isdglobal.org/digital_dispatches/trouble-with-tiktok-policy-recommendations/ [accessed 2025-01-31]

102. Minnesota attorney general's report on emerging technology and its effects on youth well-being. The Office of Minnesota Attorney General Keith Ellison. URL: https://www.lrl.mn.gov/docs/2024/mandated/240175.pdf [accessed 2025-01-31]

103. Maidment K, Tonna Z, Houlihan M, Carbone S. The impact of screen time and social media on the mental health of young Australians. Melbourne: Prevention United. URL: https://nest.greenant.net/index.php/s/QiR56KZpQzMPPBn?mc_cid=8b145254ee&mc_eid=67df8ff35d [accessed 2025-01-31]

104. Sharrock S, Hudson N, Kerr J, Chalker C, David M, Myers C. Key attributes experiences of cyberbullying among children in the UK. National Centre for Social Research. URL: https://tinyurl.com/2bt93v33 [accessed 2025-01-31]

105. Stanley L, Tanner W, Treadwell J, Blagden J. The kids aren't alright: the 4 factors driving a dangerous detachment from democracy. UK Onward. URL: https://www.ukonward.com/reports/the-kids-arent-alright-democracy/ [accessed 2025-01-31]

106. Johnson A. How to address children's online safety in the United States. Information Technology & Innovation Foundation. URL: https://www2.itif.org/2024-child-online-safety.pdf [accessed 2025-01-31]

107. Protecting the age of innocence. House of Representatives Standing Committee on Social Policy and Legal Affairs. URL: https://www.aph.gov.au/Parliamentary_Business/Committees/House/Social_Policy_and_Legal_Affairs/Onlineageverification/Report [accessed 2025-01-31]

108. Influencer culture: lights, camera, inaction. Digital, Culture, Media and Sport Committee. URL: https://publications.parliament.uk/pa/cm5802/cmselect/cmcumeds/258/report.html [accessed 2025-01-31]

109. Tackling online abuse: petitions commons select committee's inquiry on online abuse and the experience of disabled people. Inclusion London. URL: https://committees.parliament.uk/writtenevidence/9813/pdf/ [accessed 2025-01-31]

110. Social media and Australian society. Orygen Institute & Headspace. URL: https://tinyurl.com/2yemvptn [accessed 2025-01-31]

111. Best interests and targeting: implementing the privacy act review to advance children's rights. Reset Australia. URL: https://au.reset.tech/uploads/Best-Interests-Report-240128-digital.pdf [accessed 2025-01-31]

112. Galea S, Buckley G. Social media and adolescent mental health: a consensus report of the National Academies of Sciences, Engineering, and Medicine. PNAS Nexus. Mar 2024;3(2):pgae037. [FREE Full text] [doi: 10.1093/pnasnexus/pgae037] [Medline: 38415222]

113. The incelosphere: exposing pathways into incel communities and the harms they pose to women and children. Center for Countering Digital Hate Quant Lab. URL: https://counterhate.com/wp-content/uploads/2023/08/CCDH-The-Incelosphere-FINAL.pdf [accessed 2025-01-31]

114. Dunn S, Vaillancourt T, Brittain H. Supporting safer digital spaces. Centre for International Governance Innovation. URL: https://www.cigionline.org/static/documents/SaferInternet_Special_Report.pdf [accessed 2025-01-31]

115. Lenhart A, Owens K. The unseen teen - the challenges of building healthy tech for young people. Data and Society. URL: https://datasociety.net/library/the-unseen-teen/ [accessed 2025-01-31]

116. Perspective Economics, Aiken M, Davidson J. Safer technology, safer users: the UK as a world-leader in safety tech. University of East London. URL: https://assets.publishing.service.gov.uk/media/60622469e90e072d9af1df16/Safer_technology__safer_users-_The_UK_as_a_world-leader_in_Safety_Tech_V2.pdf [accessed 2025-01-31]

117. VoCO - verification of children online phase 2 report. Government Communications Headquarters; Department for Digital, Culture, Media & Sport. URL: https://assets.publishing.service.gov.uk/media/5faa9cffd3bf7f03a841cfc2/November_VoCO_report_V4__pdf.pdf [accessed 2025-01-10]

118. The impact of body image on mental and physical health. Health and Social Care Committee. URL: https://committees.parliament.uk/publications/23284/documents/170077/default/ [accessed 2025-01-10]

119. Colliver C, King J. The first 100 days: coronavirus crisis management on social media platforms. Institute for Strategic Dialogue. URL: https://www.isdglobal.org/wp-content/uploads/2020/06/First-100-Days.pdf [accessed 2025-01-10]

120. Gallagher A, Cooper L, Bhatnagar R, Gatewood C. Pulling back the curtain: an exploration of YouTube's recommendation algorithm. Institute for Strategic Dialogue. URL: https://www.isdglobal.org/wp-content/uploads/2024/06/Pulling-Back-the-Curtain-Executive-Summary-6.pdf [accessed 2025-01-31]

121. Online antisemitism: a toolkit for civil society. Institute for Strategic Dialogue. URL: https://unesdoc.unesco.org/ark:/48223/pf0000381856 [accessed 2025-01-10]

122. Social media and adolescent health. National Academies of Sciences, Engineering, and Medicine. URL: https://doi.org/10.17226/27396 [accessed 2025-01-31]

123. Responding to the social and economic drivers of youth mental health: policy lab. Orygen Institute. URL: https://tinyurl.com/ytwhar3f [accessed 2025-01-10]

124. Influence of international digital platforms. Economics References Committee. URL: https://parlinfo.aph.gov.au/parlInfo/download/committees/reportsen/RB000119/toc_pdf/Influenceofinternationaldigitalplatforms.pdf [accessed 2025-01-31]

125. Chaudhary C. Harnessing the feed: social media for mental health information and support. ReachOut. URL: https://d1robvhmkdqpun.cloudfront.net/46e49639306f108e035212644ba15d45.pdf [accessed 2025-01-31]

126. Williams D, McIntosh A, Farthing R. Profiling children for advertising: Facebook's monetisation of young people's personal data. Reset Australia. URL: https://au.reset.tech/uploads/resettechaustralia_profiling-children-for-advertising-1.pdf [accessed 2025-01-31]

127. The future of digital regulation in Australia: five policy principles for a safer digital world. Reset Australia. URL: https://au.reset.tech/uploads/the-future-of-digital-regulations-in-australia.pdf [accessed 2025-01-31]

128. Digital platform regulation green paper. Reset Australia. URL: https://au.reset.tech/uploads/Digital-Platform-Regulation-Green-Paper.pdf [accessed 2025-01-31]

129. Towards a suicide-safer internet. Samaritans. URL: https://nspa.org.uk/wp-content/uploads/2022/02/Samaritans_WhatASafeInternetLooksLike_2022.pdf [accessed 2025-01-31]

130. Unsafe children: driving up our country's response to child sexual abuse and exploitation. The Centre for Social Justice. URL: https://www.centreforsocialjustice.org.uk/wp-content/uploads/2021/03/CSJJ8804-Unsafe-Children-210325-WEB.pdf [accessed 2025-01-31]

131. Protecting children in the online world. UNICEF Australia. URL: https://assets-us-01.kc-usercontent.com/99f113b4-e5f7-00d2-23c0-c83ca2e4cfa2/6cde226b-23d1-413a-bac3-7f0eafe524d4/UA_Digital-Wellbeing-Position-Paper-2024_LR_FINAL.pdf [accessed 2025-01-31]

132. Our epidemic of loneliness and isolation. The U.S. Surgeon General's Advisory, US Department of Health and Human Services. URL: https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf [accessed 2025-01-31]

133. Teens, screens and mental health. World Health Organization. URL: https://www.who.int/europe/news-room/25-09-2024-teens--screens-and-mental-health [accessed 2025-01-31]

134. Livingstone S, Stoilova M, Stänicke L. Young people experiencing internet-related mental health difficulties: the benefits and risks of digital skills. An empirical study. ySkills. URL: https://eprints.lse.ac.uk/116407/3/D6.1_ySkills_WP6.4_Mental_health_difficulties_and_digital_skills_Report_Final.pdf [accessed 2025-05-29]

135. West D. How disinformation defined the 2024 election narrative. Brookings. URL: https://www.brookings.edu/articles/how-disinformation-defined-the-2024-election-narrative/ [accessed 2025-01-31]

136. Dufeal AL. Norway aims to raise age limit for social media users. Brussels Signal. URL: https://brusselssignal.eu/2024/10/norway-aims-to-raise-age-limit-for-social-media-users/ [accessed 2025-01-31]

137. Walsh A. TikTok stops working as US ban comes into force. BBC. URL: https://www.bbc.com/news/articles/cz6p1g54q85o [accessed 2025-01-31]

138. Gillett R, Stardust Z, Burgess J. Safety for whom? Investigating how platforms frame and perform safety and harm interventions. Social Media Soc. Dec 15, 2022;8(4):82. [doi: 10.1177/20563051221144315]

139. Paul K. Reversal of content policies at alphabet, meta, and X threaten democracy, warn experts. The Guardian. URL: https://www.theguardian.com/media/2023/dec/07/2024-elections-social-media-content-safety-policies-moderation [accessed 2025-01-31]

140. Latiff R. ByteDance's TikTok cuts hundreds of jobs in shift towards AI content moderation. Reuters. URL: https://www.reuters.com/technology/bytedance-cuts-over-700-jobs-malaysia-shift-towards-ai-moderation-sources-say-2024-10-11/ [accessed 2025-01-31]

141. Fried I. Meta's new policies open gate to hate. Axios. URL: https://www.axios.com/2025/01/09/meta-moderation-transgender-women-hate [accessed 2025-01-31]

142. Gilbert D. Meta ditches fact-checkers ahead of Trump's second term. Wired. URL: https://www.wired.com/story/meta-ditches-fact-checkers-in-favor-of-x-style-community-notes/ [accessed 2025-01-31]

143. Draper D, Neschke S. The pros and cons of social media algorithms. Bipartisan Policy Center. URL: https://bipartisanpolicy.org/download/?file=/wp-content/uploads/2023/10/BPC_Tech-Algorithm-Tradeoffs_R01.pdf [accessed 2025-01-31]

144. Addicted to bad news: how 'doomscrolling' is affecting your mental health. Ramsay Mental Health. URL: https://www.ramsaymentalhealth.com.au/en/resources-support/addicted-to-bad-news/ [accessed 2025-01-31]

145. Doomscrolling: breaking the habit. University Hospitals. URL: https://tinyurl.com/bddk6rhz [accessed 2025-01-31]

146. Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography. eSafety Commissioner. URL: https://www.esafety.gov.au/sites/default/files/2023-08/Age-verification-background-report.pdf?v=1736394644828 [accessed 2025-01-31]

147. Huddlestone J. Improving youth online safety without sacrificing privacy and speech. CATO Institute. URL: https://www.cato.org/briefing-paper/improving-youth-online-safety-without-sacrificing-privacy-speech [accessed 2025-01-31]

148. Taylor J. Social media age restrictions may push children online in secret, Australian eSafety commissioner says. The Guardian. URL: https://www.theguardian.com/australia-news/article/2024/jun/23/social-media-age-restrictions-may-push-children-online-in-secret-australia-regulator-says [accessed 2025-01-31]

149. Karp P, Touma R. Social media platforms with 'low risk of harm to children' could escape Albanese government age ban. The Guardian. URL: https://www.theguardian.com/australia-news/2024/oct/10/australia-social-media-ban-companies-low-harm-risk-michelle-rowland [accessed 2025-01-31]

150. Giardini T, Buza M. Global digital policy roundup: November 2024. Tech Policy Press. URL: https://www.techpolicy.press/global-digital-policy-roundup-november-2024/ [accessed 2025-01-31]
151. Morten CJ, Nicholas G, Viljoen S. Researcher access to social media data: lessons from clinical trial data sharing. Berkeley Tech Law J. 2024;39(1):109-204. [doi: 10.15779/Z38QF8JK81]

Abbreviations

AI: artificial intelligence
ISD: Institute for Strategic Dialogue
LGBTQ+: lesbian, gay, bisexual, transgender, and queer
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

XSL•FO
RenderX