

Social media and youth mental health: a scoping review of platform and policy recommendations

Jasleen Chhabra, Vita Pilkington, Ruben Benakovic, Michael Wilson, Louise La Sala, Zac Seidler

Submitted to: Journal of Medical Internet Research
on: February 02, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 68

 Figures 69

 Figure 1..... 70

 Figure 2..... 71

CONSORT (or other) checklists..... 72

 CONSORT (or other) checklist 0..... 73

Social media and youth mental health: a scoping review of platform and policy recommendations

Jasleen Chhabra¹ MBBSMD, PhD; Vita Pilkington¹; Ruben Benakovic¹; Michael Wilson¹; Louise La Sala¹; Zac Seidler²

¹ Centre of Youth Mental Health Orygen Youth Health Footscray AU

² Movember Foundation East Melbourne AU

Corresponding Author:

Jasleen Chhabra MBBSMD, PhD

Centre of Youth Mental Health
Orygen Youth Health
35 Poplar Road, Parkville
Footscray
AU

Abstract

Background: High rates of social media use among young people, alongside high mental ill-health within this cohort, have drawn significant public, policy, and research concern. Rapid technological advancements and changes in platform design have impeded researchers' and policymakers' empirical understandings 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 have been published, a comprehensive summary of these recommendations does not exist.

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

Methods: We searched academic (PubMed, SCOPUS, PsycInfo), and non-academic (Overton, 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, recommendations emphasise the regulation and/or design of social media products, and practices of social media companies. Peer-reviewed and grey literature from selected western contexts (Australia, Canada, UK, USA) were relevant for inclusion. Documents were published between January 2020 to September 2024.

Results: Of the identified 4,980 unique reports, 120 progressed to full-text screening, and 70 met 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. These themes offer insight into various strategies for responding to possible negative impacts of social media on young people's mental health and wellbeing.

Conclusions: This review emphasises the need for multi-pronged approaches to address the rapidly increasing presence and reach of social media platforms in young people's lives. These recommendations provide practical and tangible paths forward for governments and industry, backed by expert organisations in youth mental health and technology regulation, at a time where expert-informed guidance is sorely needed. Future efforts are needed to undertake rigorous evaluations of these proposed recommendations, while continuing to build upon the emerging peer-reviewed evidence-base that should form the foundation of policy and regulatory changes.

(JMIR Preprints 02/02/2025:72061)

DOI: <https://doi.org/10.2196/preprints.72061>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org/>

Original Manuscript

Social media and youth mental health: a scoping review of platform and policy recommendations



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 have been published, a comprehensive summary of recommendations for protecting young people's mental health and online safety does not yet exist.

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

Methods: We searched academic (PubMed, SCOPUS, PsycINFO), and non-academic (Overton, 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, recommendations emphasise the regulation and/or design of social media products, and practices of social media companies. Peer-reviewed and grey literature from selected western contexts (Australia, Canada, United Kingdom, United States) were relevant for inclusion. Documents were published between January 2020 to September 2024.

Results: Of the identified 4,980 unique reports, 120 progressed to full-text screening, and 70 met 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 emphasises the need for multi-pronged 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 organisations in youth mental health and technology regulation, at a time where expert-informed guidance is sorely needed. Rigorous evaluation of the proposed recommendations is needed, while continuing to build upon the emerging peer-reviewed evidence-base that should form the foundation of policy and regulatory changes.

Keywords: Social media; mental health; young people; recommendations; regulation; platform design; policy.

Introduction

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 policymakers, 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 seen in other countries, such as the United Kingdom, where an estimated 92% of young people are active on social media by age 12 [2], the United States, where 95% of 13- to 17-year-olds report using social media [3], and Canada, where 91% of 15- to 24-year-olds 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 wellbeing [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 marginalisation, such as LGBTQIA+ communities [9,10], and can facilitate access to information that may not be available or accessed by 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 wellbeing [12]. The social aspects of these online platforms,

such as the ability to message others and share content, have been linked to greater wellbeing, reduced depression, increased positive emotions, and decreased loneliness [13-17]. Being connected to support groups can also help young people feel validated, understood, and improve their ability to manage their mental ill-health (an umbrella term which encompasses a continuum from the most commonly occurring to the most severe and disabling mental health problems; [18]) and/or suicidal thoughts and behaviours [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 wellbeing [20]. However, such benefits are not universally recognised by others in young people's lives. In recent Australian surveys of 631 parents and 921 young people (aged 16-25), parents ranked social media as the leading concern in relation to their child's wellbeing, while young people ranked social media as the 24th most pressing concern regarding their wellbeing [21,22]. This exemplifies a generational disconnect where young people often cite important benefits of social media, while caregivers and other adults focus to a far greater extent on its potential harms.

Exposure to harmful content and interactions on social media have 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, anxiety [24], and in some cases, has been linked to suicide attempts in young people [25]. Exposure to disordered eating content, particularly in adolescent girls (e.g., [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, which 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 behaviours 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 wellbeing, including potentially leading to post-traumatic disorder [36] and acute stress disorder [37]. Additionally, 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 jeopardise young people's understanding of reality, resulting in confusion, fear, paranoia, and perpetuate harmful attitudes and beliefs [39]. For example, the spread of health-related misinformation (e.g., overstating risks associated with COVID-19 vaccination; [38]), has been associated with poor vaccine uptake and mental-ill health in young people (e.g., 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 employ a range of strategies for moderating posts to limit the proliferation of harmful content on their platforms, including human-facilitated and automated methods (e.g., artificial intelligence tools; [44]). However, there are concerns that users can easily circumvent moderation practices, given moderation tools can fail to detect harmful posts and have been outpaced by increasing use of artificial intelligence 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 organises, prioritises, and

promotes content into users' feeds based on user actions, preferences, and demographics; [48]) has also raised concern [49]. Recommender systems and highly engaging design features such as 'endless scrolling' (a listing-page design approach which loads content continuously as the user scrolls down; [50]) are employed by social media companies to keep users engaged with their respective platforms for extended periods [51]. Moreover, recent changes to recommender systems have resulted in de-prioritising content from family, friends and the accounts users follow, in favour of algorithmically curated content [52]. These curated, highly personalised feeds, which are shaped by user data, are designed to keep users on platforms as long as possible [53]. Some suggest this design feature can amplify young people's exposure to extreme viewpoints (e.g., misogyny, homophobia, racism, or eating disorder content; [54]), inappropriate advertising material (e.g., promotion of cosmetic procedures, diet supplements) and disinformation [55]. Importantly, engaging app features and curated content have been associated with sleep disturbance, compulsive behaviours (e.g., continuously refreshing social media apps), depression, and anxiety, resulting in overall poor mental health and wellbeing [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 created with demographic information indicating users were men aged 12-20, were frequently shown harmful content promoting restrictive masculine norms and anti-feminist 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 designed 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 minimise 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 (e.g., allowing legal but likely harmful content such as targeted advertisements; [67]) and the lack of international consistency in legislation [64] further complicates efforts to minimise 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 (e.g., [68,69]). One approach to mitigate negative impacts of social media on young people being pursued in some Western countries is restricting social media access. For example, the Australian Government has recently legislated restriction of young people's (<16s) access to social media from 2025 [70]. Similar measures are being discussed in both the US and the UK, with the UK considering the potential restriction of young adults' access to social media [71].

The present 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 surrounding 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 is presently without a comprehensive synthesis of recommendations. This scoping review therefore aimed to consolidate existing recommendations surrounding how social media can effectively be designed and regulated to prioritise protection of young people's mental health and minimise harm. The review synthesises available recommendations directed towards

governments, regulators, and social media companies, using thematic analysis to group them into categories reflecting similar aims or principles.

Method

Search Strategy

The review and reporting methodology followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [72]. The search strategy was prepared in consultation with a university librarian. Included documents were identified through two sources (Overton and Google) using 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 grey literature sources. Although peer-reviewed literature was eligible for inclusion (and was also considered and trialled when developing the search strategy but yielded few relevant documents), it was deemed more appropriate to prioritise grey literature documents that offered tangible recommendations to governments, regulators, and members of the social media industry, such as policy briefs, reports from non-government 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 was therefore 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 re-ran our search strategy in the following academic databases (e.g., PubMed, SCOPUS, PsycINFO), however no additional relevant results were identified.

Selection Criteria

To be included in this study, documents were required to (a) focus on young people between

ages 12-25 years, (b) present recommendations for social media companies and/or governments/regulators that promote safe social media products and services for young people, (c) be peer-reviewed published journal articles (original research), case studies, or grey literature (e.g., policy briefs, reports), (d) be published in Australia, Canada, United Kingdom, and United States or published by global organisations (e.g., Amnesty International, WHO), (e) published between January 2020 to September 2024 (due to proliferation in social media platforms and its users after COVID-19; [74]), and (f) published in English.

Data Screening

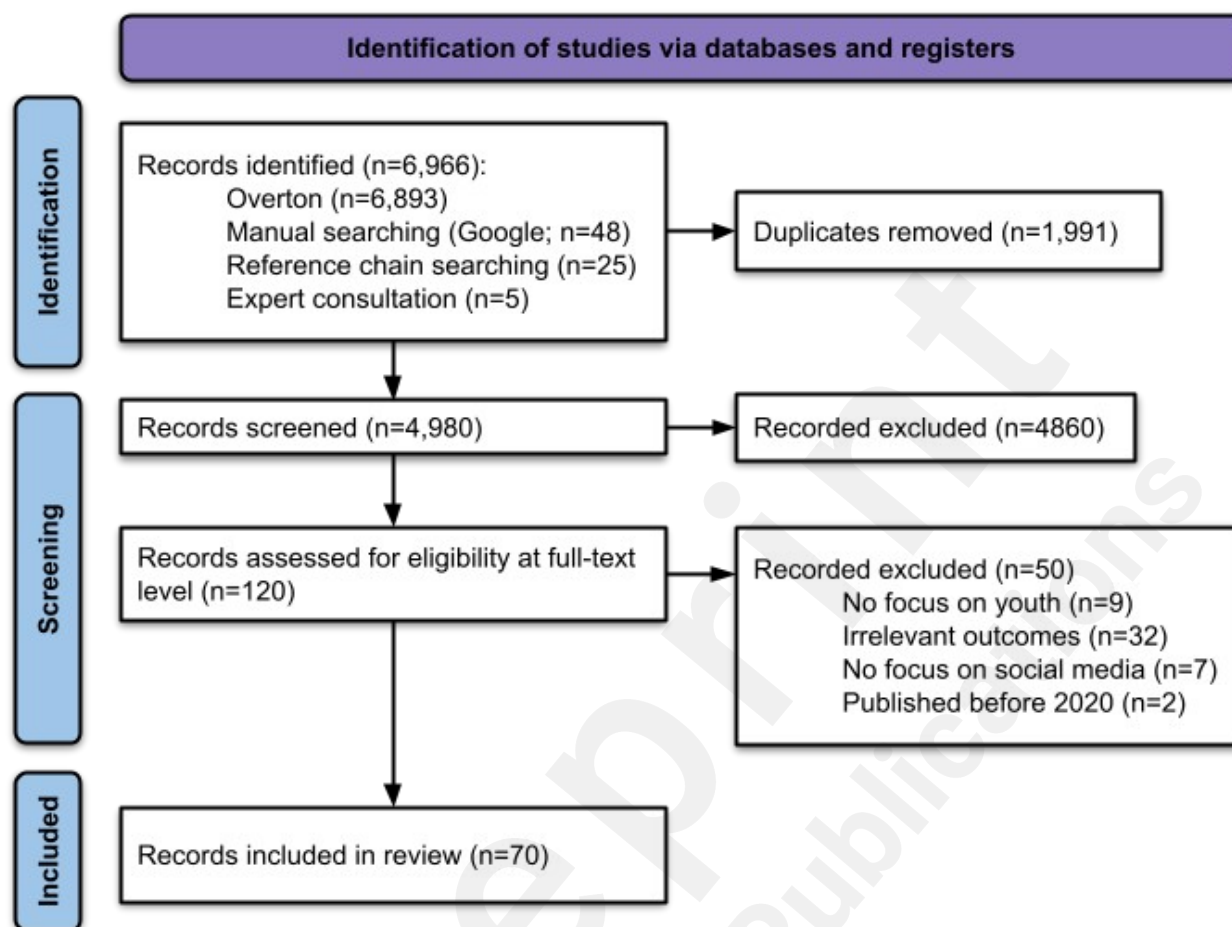
Search outputs were imported to Covidence systematic review software [75]. The first two authors independently screened all titles and abstracts against the inclusion criteria. Results were compared with a disagreement rate of less than 2% recorded (Gwet's AC1 = 0.97, Cohen's kappa = 0.54). Discrepancies were discussed until consensus was reached between the authors. For the second stage of the review, dual screening was performed on 25% of 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, author(s), organisation, year, country, type of document, target outcomes, and target audience. Information such as recommendations to social media platforms (e.g., improved platform design) and to the government bodies (e.g., strengthening existing legislations) was extracted. To collate and synthesise the recommendations provided, all documents were exported to NVivo 14 [76]. The first and second authors analysed the included recommendations, grouping recommendations into preliminary themes and sub-themes to capture conceptually similar ideas. The preliminary themes were then reviewed by the wider co-author team, and through a collaborative process, the thematic structure was refined to ensure conceptual clarity, coherence, and alignment with the study aims.

Results

The search strategy returned 6,966 reports, of which 1,911 were duplicates. In the initial screening stage, 4,980 reports were assessed for relevance. Following the exclusion of 4,860 non-relevant documents, 120 reports progressed to full-text screening, leading to the inclusion of 70 reports in the scoping review (Figure 1).

Figure 1. PRISMA Diagram

Report Characteristics

Among the 70 included reports, 24% (n=17) provided recommendations to regulators/governments, 20% (n=14) provided recommendations to social media companies, and 56% (n=39) provided recommendations to both regulators/governments and social media companies. The reports sought to target the following outcomes for young people: mental health and wellbeing related to social media use (n=35), including exposure to content related to self-harm and suicidality (n=7); exposure to misinformation and disinformation (n=16); exposure to discrimination, hate crime and hate speech (n=6), including extremist content (n=9), misogynistic content (n=5), and cyberbullying (n=3); content that increases risk of body image concerns and disordered eating (n=8); and exposure to pornography and online child sexual abuse and exploitation (n=9). Most reports

were published in Australia (31%, n=22), the United Kingdom (26%, n=18), and the United States (24%, n=17). Characteristics of the included reports are shown in Table 1.

Table 1. Characteristics of included studies

Reference #	Organisation	Individual authors (if applicable)	Year	Title	Country	Target outcome(s)
[77]	Advancing Learning and Innovation Gender Norms (ALiGN)	Stephanie Diepeveen	2024	How Does Social Media Influence Gender Norms Among Adolescent Boys?	UK	Misogyny
[78]	American Psychological Association (APA)	NA	2023	Health Advisory on Social Media Use in Adolescence	USA	Mental Health & Wellbeing
[79]	Amnesty International	NA	2023	"I Feel Exposed": Global Caught in TikTok's Surveillance Web	Global	Mental health & Wellbeing, Suicide, and Self-harm
[33]	Amnesty International	NA	2023	Driven Into the Global Darkness: How TikTok's 'For You' Feed Encourages Self-Harm and Suicidal Ideation	Global	Suicidal Ideation and Self-harm
[56]	Black Dog Institute	Kate Maston, Lyndsay Brown, & Aliza Werner-Seidler	2024	Adolescent screen use and mental health: Summary of findings from the Future Proofing Study	AUS	Mental Health & Wellbeing
[80]	Boston University	Sandro Galea & Gillian J Buckley	2024	Social media and adolescent mental health: A consensus report of the National Academies of Sciences, Engineering, and Medicine	USA	Mental Health & Wellbeing
[81]	Butterfly Foundation	NA	2024	Social Media, Body Image and Eating Disorders	AUS	Eating Disorders
[82]	Center for Countering Digital Hate	NA	2022	The Incelosphere	UK	Mental Health & Wellbeing, Self-harm & Suicide, Misogyny, Extremism, Misinformation & Disinformation, Hate Speech
[83]	Centre for Data Ethics and Innovation	NA	2020	Online targeting: Final report and recommendations	UK	Mental Health & Wellbeing

[84]	Centre for International Governance Innovation (CIGI)	Suzie Dunn, Tracy Vaillancourt & Heather Brittain	2023	Supporting Safer Digital Spaces	CANADA	Technology-facilitated gender-based violence (TFGBV)
[85]	CIGI	Samantha Bradshaw & Tracy Vaillancourt	2024	Freedom of Thought, Social Media and the Teen Brain	CANADA	Mental Health & Wellbeing
[86]	Centre on Regulation and Markets at Brookings	Sanjay Patnaik & Robert E. Litan	2023	TikTok Shows why Social Media Companies Need More Regulation	USA	Misinformation, Disinformation
[87]	Data & Society	Amanda Lenhart & Kellie Owens	2021	The Unseen Teen - The challenges of building healthy tech for young people	USA	Mental health & Wellbeing
[88]	Department for Digital, Culture, Media and Sport (DCMS)	Perspective Economics, Mary Aiken, & Julia Davidson	2020	Safer Technology, Safer Users: The UK as a World-Leader in Safety Tech	UK	Misinformation, Disinformation, Hate Crime, Online Abuse (broadly)
[89]	DCMS	NA	2022	Influencer Culture: Lights, Cameras, Inaction?	UK	Children and YP as influencers: Online Harassment, Privacy Risk, Child Abuse
[90]	Eating Disorders Families Australia	NA	2022	Social Media and Online Safety	AUS	Eating Disorders
[91]	EKO	NA	2023	Suicide, Incels, and Drugs: How TikTok's deadly algorithm harms kids	USA	Suicide & Self-Harm
[10]	eSafety Commissioner	NA	2024	Tipping the balance - LGBTIQ+ teens' experiences negotiating connection, self-expression and harm online	AUS	Mental Health & Wellbeing
[92]	Govt. Communications Headquarters & DCMS	NA	2020	VoCO - Verification of Children Online	UK	Childhood Sexual Abuse (CSA)
[93]	Health and Social Care Committee	NA	2022	The Impact of Body Image on Mental and Physical Health	UK	Body Image, Body Dysmorphia, Eating Disorders, Mental Health and Wellbeing
[94]	HOPELAB, Common Sense Media, NORC (University of Chicago)	Mary Madden, Angela Calvin, and Alexa Hasse	2024	A Double-Edged Sword: How Diverse Communities of Young People Think about the Multifaceted Relationship Between Social	USA	Mental Health and Wellbeing

					Media and Mental Health		
[95]	House of Commons	of Women and Equalities Committee	2021	Changing the Perfect Picture: An Inquiry into Body Image	UK	Body Image	
[96]	Inclusion London	NA	2023	Tackling Online Abuse	UK	Discrimination, Crime	Hate
[97]	Information Technology and Innovation Fund	Ashley Johnson	2024	How to Address Children's Online Safety in the United States	USA	CSA	
[98]	Institute for Strategic Dialogue (ISD)	Chloe Colliver & Jennie King	2020	The First 100 Days: Coronavirus and Crisis Management on Social Media Platforms	Global	Misinformation, Disinformation	
[41]	ISD	Ciarán O'Connor	2021	Hatescape: An in-depth analysis of Extremism and Hate Speech on TikTok	Global	Extremism & Speech	Hate
[39]	ISD	Mauritius Dorn, Sara Bundtzen, Christian Schwieter, & Milan Gandhi	2023	Emerging Platforms and Technologies: An Overview of the Current Threat Landscape and its Policy Implications	Global	Misinformation, Disinformation, Extremism, Online harm in general	
[99]	ISD	Sara Bundtzen	2023	Misogynistic Pathways to Radicalisation: Recommended Measures to Assess and Mitigate Online Gender-Based Violence	Global	TFGBV, Extremism, Radicalisation	Misogyny,
[100]	ISD	Ciarán O'Connor, Jared Holt, Lucy Cooper and Kevin D. Reyes	2024	TikTok Series: Policy Recommendations	Global	Extremism & Speech	Hate
[101]	ISD	Aoife Gallagher, Lucy Cooper, Rhea Bhatnagar, and Cooper Gatewood	2024	Pulling Back the Curtain: An Exploration of YouTube's Recommendation Algorithm	Global	Misinformation, Misogynistic Content (spec. Andrew Tate), Sexual Crime	
[102]	ISD, B'nai B'rith International, & UNESCO	NA	2022	Online Antisemitism: A Toolkit for Civil Society	UK	Antisemitism, Extremism, Disinformation	
[58]	ISD & Reset. Tech Australia	Elise Thomas & Kata Balint	2022	Algorithms as a Weapon Against Women: How YouTube Lures Boys and Young Men into	AUS	Misogynistic Content	

				the 'Manosphere			
[103]	JED Foundation	NA	2024	The Jed Foundation USA (JED) Recommendations for Safeguarding Youth Well-Being on Social Media Platforms	Mental Health & Wellbeing		
[104]	Kids Online Health and Safety Task Force	NA	2024	Online Health and Safety for Children and Youth: Best Practices for Families and Guidance for Industry	Mental Health & Wellbeing		
[105]	National Academies of Sciences, Engineering, and Medicine	NA	2024	Social Media and Adolescent Health	Mental Health & Wellbeing, Misinformation, & Disinformation		
[106]	National Centre for Social Research & City, University of London	Sarah Sharrock, Nathan Hudson, Jane Kerr, Charlotte Chalker, Maria David, Carrie-Anne Myers	2024	Key attributes and experiences of cyberbullying among children in the UK	Cyberbullying		
[107]	NYC Health	NA	2023	New York City's Role in the National Crisis of Social Media and Youth Mental Health	Mental Health & Wellbeing		
[108]	Office of the New York State Attorney General Letitia James	NA	2022	Investigative Report on the role of online platforms in the tragic mass shooting in Buffalo	Extremism, Hate Crime, Misinformation		
[109]	Onward	Luke Stanley, Will Tanner, Jenevieve Treadwell, James Blagden	2022	The Kids Aren't Alright: The 4 Factors Driving a Dangerous Detachment from Democracy - Onward	Mental Health & Wellbeing (spec. Loneliness)		
[110]	Open Technology Institute	Sarah Forland, Nat Meysenburg, & Erika Solis	2024	Age Verification: The Complicated Effort to Protect Youth Online	Mental Health & Wellbeing		
[111]	Orygen	NA	2024	Responding to the Social and Economic Drivers of Youth Mental Health	Mental Health & Wellbeing		
[112]	Orygen Headspace	& NA	2024	Social Media and Australian Society: Orygen & Headspace Submission	Mental Health & Wellbeing		

[113]	Parliament of the Commonwealth of Australia	House of Representatives Standing Committee on Social Policy and Legal Affairs	2020	Protecting the age of innocence	AUS	Online Pornography
[114]	Parliament of the Commonwealth of Australia	Economics References Committee	2023	Influence of International Digital Platforms	AUS	Misinformation, Disinformation, CSA
[115]	Prevention United	Maidment, K, Tonna, Z, Houlihan, M, and Carbone, S.	2024	The Impact of Screen Time and Social Media on the Mental Health of Young Australians	AUS	Mental Health & Wellbeing
[116]	Public Policy Forum	NA	2022	Canadian Commission on Democratic Expression: How to Make Online Platforms More Transparent and Accountable to Canadian Users	CANADA	Extremism, Disinformation, Misinformation
[117]	RAND Europe	Kate Cox, Theodora Ogden, Victoria Jordan, Pauline Paille	2021	COVID-19, Disinformation and Hateful Extremism	UK	Extremism
[118]	ReachOut	Camilla Chaudhary,	2024	Harnessing the Feed	AUS	Mental Health & Wellbeing
[52]	ReachOut, Beyond Blue, & Black Dog Institute	NA	2024	Inquiry into social media impacts on Australian society	AUS	Mental Health & Wellbeing
[119]	Reset. Tech Australia	Dylan Williams, Alexandra McIntosh, & Rys Farthing	2021	Profiling Children for Advertising: Facebook's Monetisation of Young People's Personal Data	AUS	Mental Health & Wellbeing, Misinformation, Pornography
[120]	Reset. Tech Australia	NA	2022	The Future of Digital Regulation in Australia: Five Policy Principles for a Safer Digital World	AUS	Mental Health & Wellbeing, Misinformation, & Disinformation
[54]	Reset. Tech Australia	Rys Farthing	2022	Designing for Disorder: Algorithms amplify pro-anorexia content to teens and children as young as 10	AUS	Eating Disorder, Body Image, Body Dysmorphia
[65]	Reset. Tech Australia	NA	2024	A duty of care in Australia's Online Safety Act	AUS	Mental Health & Wellbeing

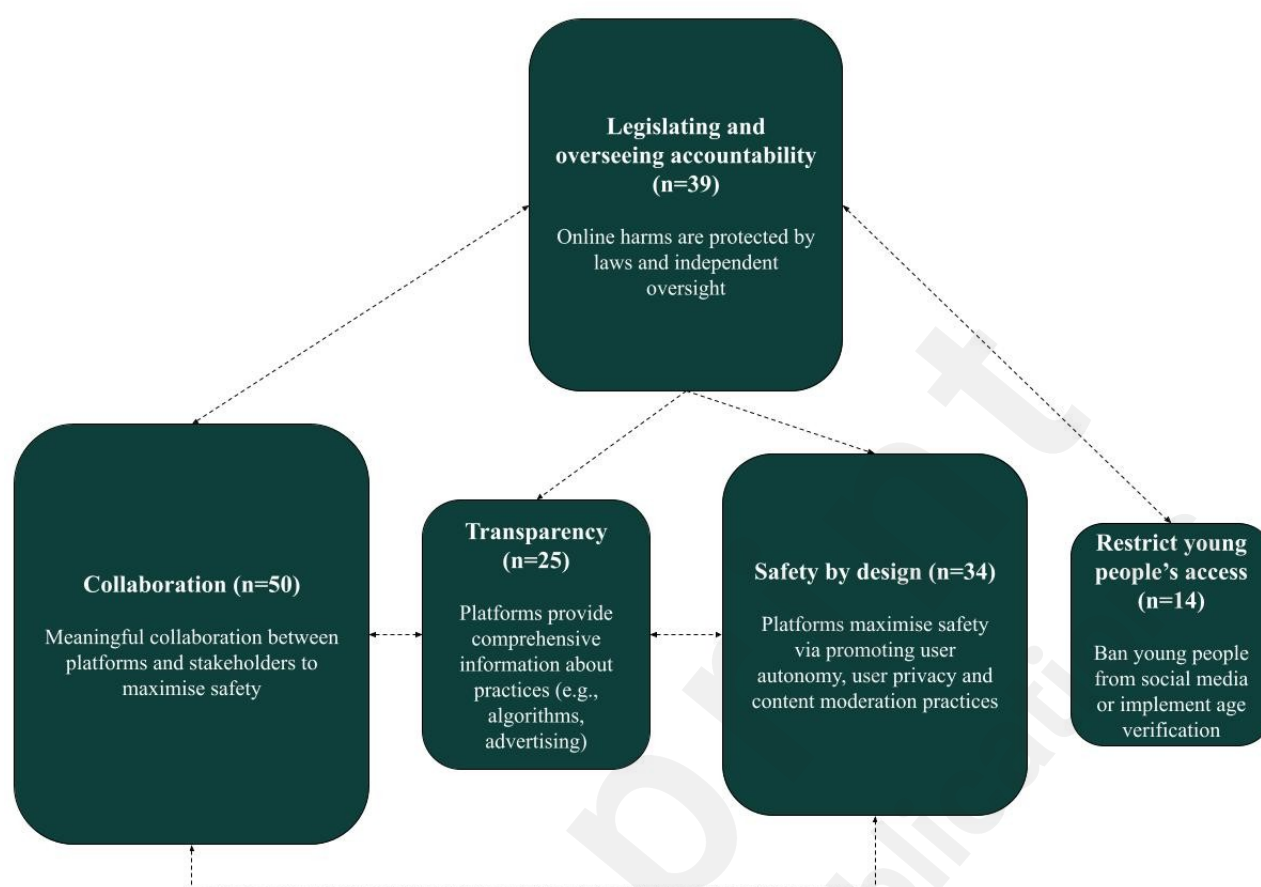
[121]	Reset. Australia	Tech	NA	2024	Achieving Platform Transparency in Australia	Digital Public in AUS	Misinformation & Disinformation
[122]	Reset. Australia	Tech	NA	2024	Best Interests & Targeting: Implementing the Privacy Act Review to advance children's rights	AUS	Mental Health & Wellbeing
[123]	Reset. Australia	Tech	NA	2024	Digital Regulation Paper	Platform Green AUS	Mental Health & Wellbeing, Misinformation, Disinformation, & CSA
[69]	Reset. Australia & Hannah Jarman	Tech	NA	2024	Not Just Algorithms: Assuring User Safety Online with Systemic Regulatory Frameworks	AUS	Eating Disorders
[124]	Reset. Australia, ChildFund Australia, and Australian Child Rights Taskforce	Tech	Rys Farthing	2022	How outdated approaches to regulation harm children and young people and why Australia urgently needs to pivot	AUS	Mental Health & Wellbeing, Misinformation, Disinformation, & CSA
[125]	Royal College of Psychiatrists (RCPSYCH)		Bernadka Dubicka & Louise Theodosiou	2020	Technology use and the mental health of children and young people	UK	Cyberbullying, Self-harm, Suicide and Body Image
[126]	Royal Society for Arts (RSA)		Asheem Singh & Jake Jooshandeh	2021	Platforms and the Public Square: A taxonomy of misinformation and the misinformed	Global	Misinformation, Disinformation
[127]	Samaritans		NA	2022	Towards a Safer Internet	UK	Suicidal Ideation and Self-harm
[31]	Scottish Government and Mental Health Foundation		NA	2020	Body Image: We are more than what we look like	UK	Body Image
[128]	The Centre for Social Justice		NA	2021	Unsafe Children: Driving up our country's response to child sexual abuse and exploitation	UK	CSA
[129]	The Office of Minnesota Attorney General Keith Ellison		NA	2024	Minnesota Attorney General's Report on Emerging Technology and its Effects on Youth Well-Being	USA	Mental Health & Wellbeing, CSA
[130]	UNICEF Australia		NA	2024	Protecting Children in the Online World	AUS	Mental Health & Wellbeing, Cyberbullying
[131]	US Surgeon		NA	2021	Protecting Youth	USA	Mental Health &

	General				Mental Health		Wellbeing	
[132]	US General	Surgeon	NA	2023	Our Epidemic of Loneliness and Isolation	USA	Depression, Anxiety, Suicidality, and Self-harm	
[3]	US General	Surgeon	NA	2023	Social Media and Youth Mental Health	USA	Mental Health & Wellbeing	
[133]	WHO		NA	2024	Teens, Screens and Mental Health	Global	Mental Health & Wellbeing	
[134]	ySKILLS		Sonia Livingstone, Mariya Stoilova, Line Indrevoll Stanicke, Reider Schei Jessen, Richard Graham, Elisabeth Staksurd, Tine Jensen	2022	Young People Experiencing Internet-Related Mental Health Difficulties: The Benefits and Risks of Digital Skills	UK	Mental Health and Wellbeing	

Synthesis of Evidence

The content analysis resulted in identification of five core interrelated themes, as shown in Figure 2. Four of the five themes centred around strategies for social media companies and regulators to facilitate the provision of safer products to young people, while one theme centred around restricting young people's access to social media altogether.

Figure 1: Thematic map of the five 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 the theme was present in a higher number of reports.

The five themes are described briefly in Table 2 and in further detail below, with illustrative quotes throughout.

Table 2. Description of themes and sub-themes

Theme 1	Legislating and overseeing accountability (n=39) Recommendations surrounding the need for governments and independent regulatory bodies to hold greater levels of responsibility and accountability to ensure (i) users are not harmed, (ii) when users are harmed, the harms are addressed appropriately, and (iii) failure to address harmful content leads to penalties for social media companies.		
Sub-themes	a. Creating or Reforming Legislation Legislative changes to facilitate ensuring social media companies are accountable in protecting user safety.	b. Expanding definition of online harms Broadening the current definitions of online harms. Included incorporating harms associated with body image and technology-facilitated gender-based violence.	c. Empowering and Equipping Regulators Strategies for ensuring regulators are empowered and equipped to hold social media companies accountable.
Theme 2	Transparency (n=25) Recommendations on the ways social media companies can facilitate access to clear information about their products and practices (e.g., with users, governments, researchers).		
Sub-	Advertising and algorithmic transparency	Transparency reports	

themes	Social media platforms to provide users and others (e.g., regulators and researchers) with clear information surrounding promotional material on users' feeds and how users' algorithms are shaped.	Social media platforms to release regular prescriptive transparency reports that offer regulators and the public with clear and detailed information about content moderation policies and practices, accuracy and error rates for automated review, number of 'take-down' orders, etc.
Theme 3	Collaboration (n=50) Recommendations suggesting social media companies to collaborate with relevant stakeholders (e.g., parents, caregivers, regulators, and researchers) at all stages of service design and delivery	
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/platforms. Includes handling user data and ensuring that potential harms to users are minimised, identified, and addressed.	
Sub-themes	a. Privacy Recommendations to enhance user privacy and data protection legislations to limit collection and use of user data by social media companies.	b. Content moderation Social media companies' roles in determining content that is appropriate and safe to appear on their platforms, and actions taken to prevent and remove harmful online content from appearing and proliferating on these platforms.
		c. Autonomy Strategies taken by social media companies (including those responsible for designing, developing, and delivering services) to increase platform users' decision-making about their (user) experiences on social media platforms.
Theme 5	Restricting young people's access to social media (n=14) Recommendations surrounding implementing age verification measures and banning young people from accessing social media platforms/services until specified ages.	

Theme 1. Legislating and overseeing accountability

The first theme *Legislating and overseeing accountability* encompasses recommendations surrounding 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 (i) users are not harmed, and (ii) when users are harmed, these are addressed appropriately and, (iii) 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 one contained three sub-themes, discussed below.

1a. Creating or reforming legislation

The sub-theme *creating or reforming legislation* appeared in 14 reports and referred to legislative changes required to ensure social media companies are held accountable for protecting young users' mental health. Many recommendations centred around making social media companies

liable for allowing dissemination of harmful user-generated content on their platforms (e.g., [116]). This included imposing penalties “sufficient to realise the goal of deterrence” [108 (p42)]. Some reports referred broadly to ‘harmful material’ posted on social media platforms, while others specifically discussed the need to impose liabilities surrounding content that encourages extreme dieting, eating disorders, self-harm [90] and violent crimes [108].

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. - Canadian Commission on Democratic Expression [116]

A number of reports recommended changes to Section 230(c)(1-2) of the 1996 Communications Decency Act in the United States (e.g., [86,108]), 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 (e.g., Meta, ByteDance) accountable for harmful posts (e.g., relating to self-harm or hate speech) posted and shared by users. Implementing such legislative changes, with associated penalties for non-compliance, would likely encourage platforms to enact more effective self-regulation [86].

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. - Office of the New York State Attorney General Letitia James [108]

1b. Expanding definitions of online harms in legislation

The sub-theme, *expanding definitions of online harms*, referred to recommendations in nine

reports (n=9) around broadening current definitions of online harms in relevant legislation (e.g., Australia's Online Safety Act). Amending this definition was suggested to ensure consistent and comprehensive prevention and response measures surrounding 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 [95] called for inclusion of appearance-based bullying in United Kingdom legislation pertaining to online harms; the Butterfly Foundation & 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 [126] suggested the Online Safety Bill include harms caused by misinformation in its remit. Further, 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].

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. - Butterfly Foundation & Daniel [81]

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 offences such as sexual assaults occur within these emerging online spaces warrants the need to review existing criminal codes, define criminal behaviour in virtual spaces, and ensure legislation exists to prosecute such offences.

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 qualify for the legal terms of rape. - Dorn et al [39]

1c. Empowering and equipping regulators

The sub-theme *empowering and equipping regulators* appeared in 30 reports, and referred to the need to implement strategies that ensure 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 [121], establishing independent regulators (including data protection regulators; [79]), and ensuring regulators have power to impose penalties when social media companies fail to comply with relevant rules and regulations (e.g., [65,116]). Within these reports, authors highlighted the importance of external oversight, in light of the current norm of self-regulation of practices and policies by social media companies (e.g., [79, 99,121]).

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. - Amnesty International [79]

Several reports recommended establishing new independent regulators focused on auditing, investigating and suggesting reforms to the policies and practices of online platforms, and enforcement (e.g., [108,116]). Authors noted the need for such regulators to have information-gathering powers and access data held by social media companies for oversight purposes (e.g., [65,116]). This also requires regulators to be appropriately informed and equipped to deal with the complex, ever-changing nature of online technologies and behaviour [83].

To meaningfully drive change, regulations need to be enforceable, and regulators must be empowered and resourced - Reset Australia [65]

Theme 2. Transparency

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

2a. Advertising and algorithmic transparency

Calls for advertising and algorithmic transparency appeared in 16 reports. These recommendations emphasised 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 surrounding 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 (e.g., [65,69,83]). It was recommended that these repositories include clear details about the advertiser, advertisement content, targeting parameters (i.e., specific demographics of users being targeted), demographics of those exposed to advertisements, and the advertising spend and length of campaigns [83,116,121]. Several reports noted a need for greater transparency surrounding political advertising, in particular (e.g., [83,116,121]). For example, during the United States 2024 elections, several reports suggested that disinformation on platforms such as Twitter and Facebook heavily influenced individuals' view of candidates and their policies on economy, immigration, and crime [135].

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. - Canadian Commission on Democratic Expression [116]

Recommendations also advocated for transparency regarding algorithms, where social media companies should articulate to users how their algorithms (or ‘feeds’) are shaped (e.g., [81,91]). Authors noted that social media algorithms operate *opaquely, with little transparency regarding how content is prioritised and served to users* [52 (p23)].

[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 - The Jed Foundation (JED) [103]

Recommendations also pertained to social media influencers and celebrities. These included calls that these figures be made subject to stricter regulation, ensuring any promotional, paid, or incentivised posts are clearly signposted to users [31]. It is important to acknowledge that currently, promotional or paid posts are often labelled as *paid*. However, the rapid increase in the number of social media influencers, their followers (young people) and the promotional materials present new challenges for advertising regulators, requiring stricter disclosure and transparency standards [95].

2b. Transparency reports

Sixteen 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 authors noted the need for clear rules regarding what should be included in these reports to facilitate *meaningful* transparency

(e.g., [79,99,117,121]). In particular, the Office of the New York State Attorney General Letitia James [108] and Reset Australia [121] offered a number of tangible recommendations surrounding what should be included within social media companies' routine transparency reports.

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. - Reset Australia, [121]

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

Theme 3. Collaboration

The third theme, *Collaboration* was documented in 50 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 (n=12), authors advocated for investment in research surrounding possible

impacts of social media on young people's mental health, wellbeing, and brain development (e.g., [52,85]). The need for research funding was highlighted in 12 reports (e.g., [78,125,131]), along with acknowledgment that conducting such research requires cross-sector cooperation and facilitating researchers' access to social media user data. Indeed, several reports (n=37) highlighted enhanced collaboration (i.e., between social media companies, researchers, and policymakers) will require wider data sharing agreements (e.g., [33,65,116]).

Ensuring that young people's experiences, agency and voices are central to policy development will lead to more relevant and effective interventions. - ReachOut, Beyond Blue and Black Dog Institute, [52]

Several recommendations emphasised the need to better understand the online experiences and needs of specific vulnerable groups, who are often at risk of online harms (e.g., [99]) and to hear from a diverse range of voices and prioritise cross-cultural collaboration (e.g., [77,107]). For example, this includes engaging with young boys who encounter misogynistic content in online spaces [77], in addition to engaging with young people of colour, LGBTQIA+ young people, and young people experiencing mental ill-health [94]. As outlined in the quote below, marginalised young people may be particularly vulnerable to online harms, however, they also may find highly affirming networks on these online spaces that *support* their wellbeing.

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... - Madden et al [94]

Some reports (n=7) highlighted the importance of communication and consistency *between*

different social media platforms, given the vast majority of users use multiple platforms (e.g., [41, 108]). This includes cross-sector research and collaboration (e.g., developing protocols for age restriction; see *Theme 5*) which should be subject to independent evaluation and auditing.

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. - O'Connor [41]

Theme 4. Safety-by-design

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

4a. Privacy

Fifteen reports highlighted the importance of social media companies having suitable processes, policies, and actions surrounding the collection and use of user data. Several reports strongly recommended enhanced and updated privacy (e.g., setting user privacy settings to 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 personalised algorithms and advertising to young people (e.g., [33,110,116,124]). Such legislation could support data minimisation (collecting the minimum amount of personal user data to deliver services), reduce

personalised content, and create special protections for particularly sensitive data (e.g., biometric information, precise location data), particularly for young users [110,124].

4b. Content moderation

Eighteen reports explored recommendations for content moderation policies, practices, and design features employed by social media companies. Recommendations advocated for stronger investment into content moderation staffing and technologies (e.g., [108]), increased use of human moderation (e.g., [33,104]), and use of advanced artificial intelligence (AI) for detecting and removing content that violates community guidelines (e.g., [99]). The moderation of content posted across an array of languages was also recommended (e.g., [99,121]).

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. - Amnesty International [33]

It was also recommended that social media companies do more to understand how users currently evade detection with current systems (e.g., using visual cues such as emojis or replacing blacklisted terms with similar-sounding words, or altering posts from their original forms; [41,100]).

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. - ISD, 2024 [41]

4c. Autonomy

A number of reports (n=20) provided recommendations pertaining to social media companies increasing the capacity for platform users to make informed decisions surrounding their experiences on the platform. Recommendations advocated for safety design features (e.g., opt-out; 129]) allowing

young people to be active agents in personalising their feed or algorithm, including their increased capacity to “suppress” types of content they do not wish to see on their feeds (e.g., [33,56,57,103,115]). This design feature may allow users to restrict or limit posts that negatively impact their mental health (e.g., disordered eating, misogynistic content, self-harm) on their feeds [79,131]. Offering young people greater choice about the type of content shown on their feeds is considered a potential protective factor against mental-ill health (e.g., [103,104]).

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 - The U.S. Surgeon General’s Advisory [131]

Recommendations surrounding specific design features on social media platforms were also offered. For example, allowing users greater choice and control over design features (e.g., turning off or disabling the ‘infinite scroll’ feature on TikTok; [52,104,110]) and greater choice on how their collected data is used (e.g., providing informed consent; [79,116]). This enhanced user choice may allow young people to use social media platforms in ways that better meet their needs and preferences [110]. Others offered stronger recommendations. For example, The Office of Minnesota Attorney General Keith Ellison [129] proposed that “dark patterns” (p. 6) within platform design that encourage excessive use for young people “beyond their explicit desires” - such as infinite scroll, auto-play videos, engagement-driven algorithms, and frequent notifications - be banned altogether.

Theme 5. Restricting young people’s access to social media

Several reports (n=14) discussed the potential benefits and drawbacks of implementing age restrictions on young people from accessing social media platforms until specified ages. Some authors recommended the introduction of mandatory age verification for all social media accounts to ensure that users met minimum age criteria (e.g., [95,106,109]). However, others 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 [33,110,124]. For example, introducing age verification and limiting young people's access to social media, may pose further risks to young people [110]. 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 presenting privacy implications.

While more efforts 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. - Forland et al [110]

Thus, recommendations emphasised the necessity for multi-pronged solutions, given “age verification is no substitute for privacy protections and increased user transparency and control” [110]. Reports also called for evaluation of the effectiveness of age verification technologies, as these necessarily mediate any widespread age-based ban (e.g., [104,110]). In addition, recommendations for alternative approaches to age verification via government IDs were suggested to circumvent privacy concerns [97]. For example, the Australian Government suggested taking advantage of an existing service, the Australian Digital Identity program, to facilitate age verification in ways that minimise privacy concerns [113]. Other recommendations included involving caregivers in verifying children's ages (e.g., [106,109]), using facial recognition technologies (e.g., Yoti; [97]), and collaboration between a range of stakeholders to develop best-practices and effective age-verification systems to minimise impact on young users [110,113,131].

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. -

Forland et al [110]

Finally, recommendations were made to explore alternatives to age verification processes (e.g., stronger legislation, empowered regulators, and safety-and-security-by-design features; [97,106,110;113]) to mitigate the additional privacy and safety concerns.

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. - Johnson [97]

Discussion

To our knowledge, this is the first review to synthesise recommendations surrounding 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 surrounding young people’s access to social media [70, 136], and growing tension between governments and social media companies [137]. The 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 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 (e.g., [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 highlighted that existing strategies are insufficient to prevent young people from being exposed to online harms and advocated strongly for improved content moderation practices and policies (e.g., [30,33,106]). This aligns with empirical

evidence that indicates more than three in four young people report exposure to online hate content [4,23], and findings indicating young people are approximately four times more likely to see posts related to self-harm or suicide, relative to those older than 25 years [11]. Perspectives on moderation strategies were mixed, with some advocating for increased investment in human moderators [33,104], while others 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 (e.g., Meta's decision to replace paid moderators with 'voluntary community notes' on its platforms [139-142]). However, as noted throughout the literature, over-reliance on AI-facilitated moderation and natural language processing techniques warrants careful consideration [44]. AI moderation tools often fail to recognise and remove iterations of banned material (e.g., terms like 'unaliving' or hashtags like #suwerslide instead of #suicide) [91]. 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 online material, such as posts related to self-harm or suicide. This is a considerable oversight, as previous studies have shown some young people actively seek information and support regarding these challenges online, particularly given 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 emphasised across reports, as has been advocated for widely in the extant academic literature (e.g., [44,46]). Such action is likely to garner support and foster trust in social media services among the public, based on data indicating the general community is supportive of governments and social media companies taking proactive steps to remove harmful online 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 online [11].

Recommendations also emphasised that social media companies must be held responsible for

young users' exposure to harmful online content through legislative enforcement of compliance. Strong support was expressed for legislative changes to support ensuring liability of platforms for harmful content being created and disseminated on their platforms, rather than allowing immunity (e.g., [108,116]). Alongside amending and introducing necessary legislation and expanding the definition of online harms [86,89,96], authors recommended empowering independent regulators with meaningful auditing, investigative, and enforcement powers [65, 116].

Crucially, regulations and moderation practices must be fit-for-purpose and informed by a thorough understanding of online technologies (including routinely used strategies for non-detection). They should be continually reviewed and adapted to ensure they are responsive to technological advancements and emerging online safety issues, such as use of 'bot accounts', AI-generated content, and encoded file formats, and responsive to changing online landscapes to avoid the "legal lag" [65 (p9)].

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 [56,104]. Given their developmental stage, these features are considered particularly harmful to young people [143] and have been associated with feelings of depression, anxiety, hopelessness, and isolation [144,145]. As a result, reports recommended social media platforms employ human-centred design approaches, where users (including young people) and other stakeholders (e.g., caregivers) are involved in platform design [104].

Implementing age verification in social media platforms has been a topic of considerable debate, with Australia recently mandating age restrictions for platforms like Instagram, TikTok, and Snapchat [10,70]. While some argued limiting young people's access to social media is crucial for preventing their exposure to inappropriate content (e.g., [3,146]), others expressed concerns about

feasibility and possible risks age verification may present (e.g., [110]). 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 required users to provide identification documents to sign up to online 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 [104].

Age restriction represents one of many possible responses to the perceived risks to young people's mental health that social media can present. 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 marginalised groups such as LGBTQIA+ 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 online 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 social media products are made safer for all users, including young people. In sum, while some view age restriction as a necessary step towards 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 greater balance between safeguarding youth and preserving their access to supportive online spaces [112,147].

Strategies for increasing transparency and collaboration

The need for greater transparency by social media companies was frequently highlighted in the recommendations. This included communicating openly and clearly with governments, regulators, users, and the general public about their policies and practices around the type of data that is collected from users and how it is used, how algorithms are shaped, what posts constitute paid/promotional content, and the types and frequencies of posts being removed from these online spaces (e.g., [112]).

Although several reports suggested social media companies release regular transparency reports (e.g., [41,83,117]), this recommendation has since been implemented in countries such as Australia, United States, 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 [121]. However, regulators have raised concerns over the quality and utility of current transparency reports [99]. For example, when figures about violative content are presented (e.g., 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 (e.g. posts inciting violence, posts proven to be false after fact-checking; [121]). They also do not indicate how many reported posts are *not* removed, or maximum time until removal and scope that 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 [121,150].

Transparency reports have also been criticised for being developed by those with a vested interest in their findings (i.e., the social media companies themselves) and social media companies' refusal to grant data access to researchers to independently verify the claims presented in these reports. Indeed, authors of the included reports argued that providing approved researchers access to

the data collected by social media platforms will enable evaluation and auditing of content moderation systems, platform policies, and their alignment with policies and regulations (e.g., [83,110, 121]). This echoes similar calls by academics (e.g., [151]). Researcher access may also enable auditing of actions taken by social media companies against harmful content and users who violate platform policies [78]. To minimise privacy risks and maximise user safety, access to public data should only be made available to ‘vetted’ researchers (e.g., [121,151]) and will require data sharing agreements, in addition to coordination and collaboration between social media companies and relevant regulators [104].

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 social media use is *positively* associated with important correlates of wellbeing, such as loneliness (e.g., [13]). As highlighted in previous reviews, long-term studies are essential to comprehensively assess how particular social media content and behaviours impact youth mental health, development, and behaviour over time [7,9]. Academics also increasingly emphasise the importance of considering nuances in *how* young people use social media (e.g., which platforms, types of content viewed) and for what purposes (e.g., entertainment, information gathering, 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]. Further, there is a need to consider whether particular groups are at elevated risk of being impacted by potential online harms (e.g., mental ill-health, exploitation, disinformation), as this may guide targeted prevention and intervention efforts. Providing researchers with access to real-time data, alongside substantial investment in research funding, would greatly enhance the ability to conduct such studies, enabling the monitoring of changes and trends in young people’s social media behaviour and broader effects over extended periods [78]. Importantly, researchers also recommend that empirical findings be coupled with young people’s insights and perspectives, which can be

supported via qualitative studies and meaningful consultation [6,11]. These approaches would offer more informed conclusions and evidence-based recommendations for managing impacts of social media on young people [83].

While collaboration between regulators, social media companies, and researchers is a crucial step towards creating safer platforms, young people should remain at the centre of this collaboration. Several reports recommended consulting and collaborating with the young people, including those who are typically under-represented during the development of legislation, regulation, and social media platform design (e.g., [33,39,122]). 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. Their input is therefore invaluable in identifying and addressing these challenges. By incorporating their perspectives, policies can be better equipped to tackle the complexities of online behaviour, mental health concerns, and digital safety, offering more nuanced and practical solutions that align with young people's lived experiences [52,122].

Limitations

The current review findings should be viewed in light of several limitations including the paucity of non-English documents. This limits the generalisability of results to non-English-speaking contexts and may not account for the impact of social media practice and policy youth mental health globally. While this review focused on countries that have been leading in terms of regulatory efforts and public discourse surrounding social media and youth mental health, it is suggested that future international reviews broaden their review scope and capture recommendations for governments/regulators and 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 (e.g., lack of resources to circumvent targeted

advertising and marketing). Additionally, although this is the first review to collate recommendations surrounding social media and youth mental health, the recommendations were not categorised 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 sensitising recommendations based on user age, as blanket policies that attempt to address the unique needs of both 12-year-olds and 25-year-olds under the same regulatory framework may prove ineffective.

This review predominantly captured grey literature, which precluded evaluation of the quality of reports and contributed to the widespread lack of evidence-based recommendations. Nonetheless, the recommendations were rationally derived by subject matter experts and organisations. Our focus on the available (largely grey) literature represents a pragmatic attempt to synthesise recommendations that are likely to be viewed by governments and regulators in the absence of corresponding and contemporaneous academic literature. Further, 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 this may have meant we overlooked some important recommendations for protecting young people's mental health and safety online, 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 this reduces some relevant recommendations surrounding 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, recommendations in this review were rarely informed by rigorous peer-reviewed evidence, which requires amelioration as the evidence-base proliferates. This limits current understanding surrounding the efficacy and impacts of enacting the types of recommendations

proposed within 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.

Conclusion

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 five interrelated themes highlighted within this review emphasises the range of options available to both industry and government, and the imperative to undertake a multi-pronged approach. These five themes, while distinct, are clearly interrelated and collectively offer a comprehensive array of strategies to address the increasing presence of social media companies in young people's lives, and likely their health. At a time where the empirical evidence for the harms and benefits of social media in young people is still emerging, the expert-based recommendations within 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 policymakers 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 supporting digital environments.

Conflicts of Interest

None declared.

Author Contributions

Author JC was responsible for project administration. JC and VP led the project conceptualisation,

data curation (including developing the search strategy and inclusion/exclusion criteria), investigation, formal analysis, writing – original draft, and writing – reviewing & editing. RB and MW contributed through writing – review & editing and visualisation. LLS contributed through writing – review & editing. ZS contributed through supervision and writing – review & editing.

Abbreviations

AI = Artificial intelligence

APA = American Psychological Association

CIGI = Centre for International Governance Innovation

COVID-19 = Coronavirus disease

DCMS = Department for Digital, Culture, Media, and Sport

ISD = Institute for Strategic Dialogue

LGBTQIA+ = An umbrella term referring to lesbian, gay, bisexual, transgender, queer, questioning, intersex, asexual people. The ‘+’ symbol represents other identities not specifically listed.

UK = United Kingdom

US = United States of America

References

1. Edwards B, Doery K, Arnup J, Chowdhury I, Edwards D, Hillman K. GENERATION Survey: Young People, Release 2 (Wave 1-2) [dataset]. ADA Dataverse; V5; 2023. Accessed 31 January 31. <https://doi.org/10.26193/YMMO4L>
2. Ofcom. *Children and parents: Media use and attitudes report*. Published 2024. Accessed January 31, 2025. <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>
3. The U.S. Surgeon General's Advisory. *Social media and youth mental health*. Published 2023. Accessed January 31, 2025. <https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf>
4. Statistics Canada. Young people and exposure to harmful online content in 2022. Published 2024. Accessed January 31, 2025. <https://www150.statcan.gc.ca/n1/pub/11-627-m/11-627-m2024005-eng.htm>
5. McGorry PD, Mei C, Dalal N, et al. The Lancet Psychiatry Commission on youth mental health. *Lancet Psychiatry*. 2024;11(9):731-774. doi:10.1016/S2215-0366(24)00163-9
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*. 2022;17(3):662-679. doi:10.1177/17456916211014189
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*. 2023;15(8):e42990. doi:10.7759/cureus.42990
8. Best P, Manktelow R, Taylor B. Online communication, social media, and adolescent wellbeing: A systematic narrative review. *Child Youth Serv Rev*. 2014;41:27-36.

doi:10.1016/j.childyouth.2014.03.001

9. Berger MN, Taba M, Marino JL, Lim MSC, Skinner SR. Social media use and health and well-being of lesbian, gay, bisexual, transgender, and queer youth: Systematic review. *J Med Internet Res*. 2022;24(9):e38449. doi:10.2196/38449
10. eSafety Commissioner. *Tipping the balance. LGBTQI+ teens' experiences negotiating connection, self-expression and harm online*. Aussie Kids Online. Published June 2024. Accessed January 31, 2025. <https://www.esafety.gov.au/sites/default/files/2024-06/Tipping-the-balance-June-2024.pdf>
11. Robinson J, La Sala L, Kenny B, et al. How do Australian social media users experience self-harm and suicide-related content? A national cross-sectional survey comparing young people and adults. *PsyArXiv*. Published January 2024. doi:10.31234/osf.io/5rh93
12. American Psychological Association (APA). Dictionary of psychology - mental health. Published 2018. Accessed January 31, 2025. <https://dictionary.apa.org/mental-health>
13. Dissing A, Hulvej Rod N, Gerds T, 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. PMID: 32536319. doi: 10.1177/1403494820920418.
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*. 2024;96(7):1555-1568. doi:10.1002/jad.12362
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*. 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.* 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.* 2011;4(3):207-217. doi:10.1093/phe/phr030
19. Kruzan KP, Nesi J, Hamilton JL, et al. Media influences on self-harm, suicidality, and suicide. In: Christakis DA, Hale L, eds. *Handbook of Children and Screens.* Springer, Cham; 2025. doi:10.1007/978-3-031-69362-5_20
20. Hall JA. Ten myths about the effect of social media use on well-being. *J Med Internet Res.* 2024;26(1):e59585. doi:10.2196/59585
21. ReachOut Australia. *What Parents Worry About: Carer Concerns for Youth Mental Health and Wellbeing.* Published 2024. Accessed 31 January 2025. <https://d1robvhmkdqpun.cloudfront.net/20ec30143a2719d6d3fda2a71980ea63.pdf>
22. ReachOut Australia. *What are you worried about?: Young people's stress burden and its impacts on their wellbeing.* Published 2023. Accessed January 31, 2025. <https://d1robvhmkdqpun.cloudfront.net/6ee09aefcd82ab783758e1da8dac6dce.pdf>
23. Enock 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 Electron J.* 2023. 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.* 2023;20(3):2383. doi:10.3390/ijerph20032383
25. Dyer C. Social media content contributed to teenager's death "in more than a minimal way," says coroner. *BMJ (Clin Res Ed).* 2022;379:o2374. doi:10.1136/bmj.o2374
26. Choukas-Bradley S, Roberts SR, Maheux AJ, et al. 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*. 2022;25:681-701. doi:10.1007/s10567-022-00404-5

27. Wells G, Horwitz, Seetharaman D. Facebook knows Instagram is toxic for teen girls, company documents show. *Wall Street Journal*. Published September 2021. Accessed January 31, 2025. www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girlscompany-documents-show-11631620739

28. Thai H, Davis CG, Mahboob W, et al. Reducing social media use improves appearance and weight esteem in youth with emotional distress. *Psychol Pop Media*. 2023. doi:10.1037/ppm0000460

29. Griffiths S, Harris EA, Whitehead G, 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*. 2024;51:101807. doi:10.1016/j.bodyim.2024.101807

30. Lonergan AR, Bussey K, Fardouly J, 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*. 2020;53(5):485-496. doi:10.1002/eat.23256

31. Scottish Government's Body Image Advisory Group on Good Body Image & Mental Health Foundation. *Body image: We are more than what we look like*. Published 2020. Accessed January 31, 2025. <https://www.mentalhealth.org.uk/sites/default/files/2022-06/MHF-Scotland-Body-image-report.pdf>

32. eSafety Commissioner. Digital lives of Aussie teens. Published 2021. Accessed January 31, 2025. <https://www.esafety.gov.au/research/digital-lives-of-aussie-teens>

33. Amnesty International. Driven into Darkness: *How TikTok's 'For You' Feed Encourages Self-Harm and Suicidal Ideation*. Published 2023. Accessed January 31, 2025. <https://www.amnesty.org/en/documents/POL40/7350/2023/en/>

34. Finkelhor D, Turner H, Colburn D. Prevalence of online sexual offenses against children in the US. *JAMA Netw Open*. 2022;5(10):e2234471. doi:10.1001/jamanetworkopen.2022.34471

35. Finkelhor D, Turner HA, Colburn D. Which dynamics make online child sexual abuse and

- cyberstalking more emotionally impactful: Perpetrator identity and images? *Child Abuse Negl.* 2023;137:106020. doi:10.1016/j.chiabu.2023.106020
36. Paat YF, Markham C. Digital crime, trauma, and abuse: Internet safety and cyber risks for adolescents and emerging adults in the 21st century. *Soc Work Ment Health.* 2020;19(1):18-40. doi:10.1080/15332985.2020.1845281
37. Gewirtz-Meydan A, Mitchell KJ, O'Brien JE. Sexual posttraumatic stress among investigators of child sexual abuse material. *Policing: J Policy Pract.* 2023;17:paad052. doi:10.1093/police/paad052
38. World Health Organization [WHO]. Infodemics and misinformation negatively affect people's health behaviours: New WHO review finds. Published September 1, 2022. Accessed January 31, 2025. <https://www.who.int/europe/news/item/01-09-2022-infodemics-and-misinformation-negatively-affect-people-s-health-behaviours--new-who-review-finds>
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 [ISD]. Published September 12, 2023. Accessed January 31, 2025. <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>
40. Borges do Nascimento IJ, Pizarro AB, Almeida JM, et al. Infodemics and health misinformation: a systematic review of reviews. *Bull World Health Organ.* 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 [ISD]; 2021. https://www.isdglobal.org/wp-content/uploads/2021/08/HateScape_v5.pdf
42. Johnson NF, Leahy R, Restrepo NJ, et al. Hidden resilience and adaptive dynamics of the global online hate ecology. *Nature.* 2019;573(7773):261-265. doi: 10.1038/s41586-019-1494-7

43. O'Donnell C, Shor E. "This is a political movement, friend": Why "incels" support violence. *Br J Sociol.* 2022;73(2):336-351. doi:10.1111/1468-4446.12923
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.* 2022;12:129. doi:10.1007/s13278-022-00951-3
45. Engstrom E, Feamster N. The limits of filtering: A look at the functionality and shortcomings of content detection tools. *Engine.* Published March 2017. Accessed 31 January 2025. <https://www.engine.is/the-limits-of-filtering>
46. Gerrard Y. Beyond the hashtag: Circumventing content moderation on social media. *New Media Soc.* 2018;20(12):4492-4511. doi:10.1177/1461444818776611
47. Stewart E. Detecting fake news: Two problems for content moderation. *Philos Technol.* 2021;34:923-940. doi:10.1007/s13347-021-00442-x
48. eSafety Commissioner. *Position paper - Recommender systems and algorithms.* Published December 2022. Accessed 31 January 2025. <https://www.esafety.gov.au/sites/default/files/2022-12/Position%20statement%20-%20Recommender%20systems%20and%20algorithms.pdf>
49. Metzler H, Garcia D. Social drivers and algorithmic mechanisms on digital media. *Perspect Psychol Sci.* 2024;19(5):735-748. doi:10.1177/17456916231185057
50. Interaction Design Foundation - IxDF. *What is infinite scrolling?.* Published 2023. Accessed 31 January 2025. <https://www.interaction-design.org/literature/topics/infinite-scrolling>
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.* 2019;16(14):2612. doi:10.3390/ijerph16142612
52. ReachOut, Beyond Blue, & Black Dog Institute. *Inquiry into social media impacts on Australian Society. Joint Select Committee on Social Media and Australian Society Submission 168.*

Published July 2024. Accessed 31 January 2025.

https://www.blackdoginstitute.org.au/wp-content/uploads/2024/07/Sub168_ReachOut-Beyonce-Blue-BDI.pdf

53. Yang C, Xu X, Nunes BP, Siqueira SWM. Bubbles bursting: Investigating and measuring the personalisation of social media searches. *Telematics Inform.* 2023;82:101999. doi:10.1016/j.tele.2023.101999

54. Farthing. *Designing for disorder: Algorithms amplify pro-anorexia content to teens and children as young*. Reset Australia. Published 2022a. Accessed 31 January 2025. <https://au.reset.tech/uploads/insta-pro-eating-disorder-bubble-april-22-1.pdf>

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

56. Black Dog Institute. *Adolescent screen use and mental health: Summary of findings from the Future Proofing Study*. Sydney, AU: Black Dog Institute. Published July 2024. Updated October 2024. Accessed January 10, 2025. <https://www.blackdoginstitute.org.au/research-projects/teens-screens-adolescent-mental-health-screen-use/>

57. ReachOut Australia. *Teenagers and social media*. Published 2024. Accessed January 31, 2025. <https://parents.au.reachout.com/staying-safe-online/social-media/teenagers-and-social-media>

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. Published 2022. Accessed January 31, 2025. <https://www.isdglobal.org/wp-content/uploads/2022/04/Algorithms-as-a-weapon-against-women-ISD-RESET.pdf>

59. Australian Government. *Online Safety Act 2021: Australian Government*. Published 2021. Accessed January 31, 2025. <https://www.legislation.gov.au/Details/C2021A00076>

60. Government UK. *A guide to the Online Safety Bill: United Kingdom Government*. Published 2022. Accessed January 31, 2025. <https://www.gov.uk/guidance/a-guide-to-the-online-safety-bill>

61. US Government. *Children's Online Privacy Protection Rule ("COPPA")*. Published 1998. Accessed January 31, 2025. <https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa>.
62. United States Government. *S.3663 — 117th Congress (2021-2022)*. Published 2022. Accessed January 31, 2025. <https://www.congress.gov/bill/117th-congress/senate-bill/3663/text>.
63. Government of Canada. *Implementing the Online News Act*. Published 2024. Accessed January 31, 2025. <https://crtc.gc.ca/eng/industr/info.htm>.
64. Robinson J, Thorn P, McKay S, 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 Psychiatr*. 2023;2:1274263. doi: 10.3389/frcha.2023.1274263
65. Reset Australia. *A duty of care in Australia's Online Safety Act*. Published 2024. Accessed January 31, 2025. <https://au.reset.tech/uploads/Duty-of-Care-Report-Reset.Tech.pdf>
66. La Sala L, Sabo A, Michail M, et al. Online safety when considering self-harm and suicide-related content: Qualitative focus group study with young people, policymakers, and social media industry professionals. *J Med Internet Res*. (in press). doi:10.2196/66321
67. National Society for the Prevention of Cruelty to Children [NSPCC]. *Children's experiences of legal but harmful content online*. Published 2022. Accessed January 31, 2025. <https://learning.nspcc.org.uk/media/pi1hqvez/legal-but-harmful-content-online-helplines-insight-briefing.pdf>
68. Council of Canadian Academies (CCA). *Vulnerable connections: Expert Panel on Public Safety in the Digital Age*. Ottawa (ON): Published 2023. Accessed January 31, 2025. https://www.cca-reports.ca/wp-content/uploads/2023/04/Vulnerable-Connections_FINAL_DIGITAL_EN_UPDATED.pdf
69. Reset Australia & Jarman, H. *Not just algorithms: Assuring user safety online with systemic*

regulatory frameworks. Published 2024. Accessed January 31, 2025.

<https://apo.org.au/sites/default/files/resource-files/2024-03/apo-nid326122.pdf>

70. Prime Minister of Australia. *Social media reforms to protect our kids online pass Parliament*.

Published 2024. Accessed January 31, 2025. <https://www.pm.gov.au/media/social-media-reforms-protect-our-kids-online-pass-parliament>

71. BBC News. Social media ban for under-16s 'on the table' in UK. Published November 20 2024. Accessed January 31, 2025. <https://www.bbc.com/news/articles/ce9gpdrx829o.amp>

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*. 2021;372:n71. doi: 10.1136/bmj.n71

73. Overton. Our mission. Accessed January 31 2025. <https://www.overton.io/about>

74. Gupta K, DSilva MH. Proliferation of social media during the COVID-19 pandemic: a statistical enquiry. *J Xi'an Univ Archit Technol*. 2020;12:1752-9.

75. Covidence. Accessed January 31 2025. www.covidence.org

76. Lumivero. Accessed January 31 2025. www.lumivero.com

77. 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)*.

Published February 2024. Accessed January 31, 2025.

<https://www.alignplatform.org/sites/default/files/2024-02/align-socialmedia-briefingnote-jan24-proof05.pdf>

78. American Psychological Association. Health advisory on social media use in adolescence.

Published May 2023. Accessed January 31, 2025.

<https://www.apa.org/topics/social-media-internet/health-advisory-adolescent-social-media-use.pdf>

79. Amnesty International. *"I feel exposed": Caught in TikTok's surveillance web*. Published

November 7, 2023. Accessed January 31, 2025.

<https://www.amnesty.org/en/documents/pol40/7349/2023/en/>

80. Galea S, Buckley GJ. Social media and adolescent mental health: A consensus report of the National Academies of Sciences, Engineering, and Medicine. *PNAS Nexus*. 2024;3(2):pgae037. doi:10.1093/pnasnexus/pgae037

81. Butterfly Foundation & Daniel, Z. *Social media, body image, and eating disorders*. Published May 2024. Accessed January 31, 2025.

<https://butterfly.org.au/wp-content/uploads/2024/05/Roundtable-Recommendations-FINAL-ONLINE.pdf>

82. Center for Countering Digital Hate Quant Lab. *The incelosphere: Exposing pathways into incel communities and the harms they pose to women and children*. Published September 2022. Accessed January 10, 2025. <https://counterhate.com/wp-content/uploads/2023/08/CCDH-The-Incelosphere-FINAL.pdf>

83. Centre for Data Ethics and Innovation. *Online targeting: Final report and recommendations*. Published February 2020. Accessed January 31, 2025. <https://www.gov.uk/government/publications/cdei-review-of-online-targeting/online-targeting-final-report-and-recommendations#introduction>

84. Dunn S, Vaillancourt T, Brittain H. *Supporting safer digital spaces*. Centre for International Governance Innovation. Published 2023. Accessed January 10, 2025. https://www.cigionline.org/static/documents/SaferInternet_Special_Report.pdf

85. Bradshaw K, Vaillancourt, T. *Freedom of Thought, Social Media and the Teen Brain*. Centre for International Governance Innovation. Published 2024. Accessed January 31, 2025. https://www.cigionline.org/static/documents/FoT_PB_no.9.pdf

86. Patnaik S, Litan RE. *TikTok shows why social media companies need more regulation*. Center on Regulation & Markets at BROOKINGS. Published 2023. Accessed January 31, 2025. https://www.brookings.edu/wp-content/uploads/2023/05/20230511_CRM_PatnaikLitan_TikTok_FINAL.pdf

87. Lenhart A, Owens K. The unseen teen - the challenges of building healthy tech for young people. *Data and Society*; Published May 2021. Accessed January 31, 2025. <https://datasociety.net/library/the-unseen-teen/>
88. Perspective Economics, Aiken M, Davidson J. *Safer technology, safer users: The UK as a world-leader in Safety Tech*. University of East London. Published 2020. Accessed January 10, 2025. https://assets.publishing.service.gov.uk/media/60622469e90e072d9af1df16/Safer_technology_safer_users- The UK as a world-leader in Safety Tech V2.pdf
89. Digital, Culture, Media and Sport Committee. *Influencer culture: Lights, camera, inaction?* House of Commons Committee. Published 2022. Accessed January 10, 2025. <https://publications.parliament.uk/pa/cm5802/cmselect/cmcumeds/258/report.html>
90. Eating Disorders Families Australia. *Inquiry into Social Media and Online Safety Submission* 37. Published 2022. Accessed January 31, 2025. https://www.aph.gov.au/Parliamentary_Business/Committees/House/Former_Committees/Social_Media_and_Online_Safety/SocialMediaandSafety/Submissions
91. Ekō. *Suicide, Incels, and Drugs: How TikTok's Deadly Algorithm Harms Kids*. Published 2023. Accessed January 31, 2025. https://s3.amazonaws.com/s3.sumofus.org/images/eko_Tiktok-Report_FINAL.pdf
92. Government Communications Headquarters; Department for Digital, Culture, Media & Sport. *VoCO - Verification of Children Online Phase 2 Report*. Published 2020. Accessed January 10, 2025. https://assets.publishing.service.gov.uk/media/5faa9cffd3bf7f03a841cfc2/November_VoCO_report_V4_pdf.pdf
93. Health and Social Care Committee. *The impact of body image on mental and physical health*. House of Commons. Published 2022. Accessed January 10, 2025. <https://committees.parliament.uk/publications/23284/documents/170077/default/>
94. 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. Published 2024. Accessed January 31, 2025.

https://www.commonsensemedia.org/sites/default/files/research/report/2024-double-edged-sword-hopelab-report_final-release-for-web-v2.pdf

95. Women and Equalities Committee. *Changing the Perfect Picture: An Inquiry into Body Image.* House of Commons. Published 2021. Accessed January 31, 2025.

<https://dera.ioe.ac.uk/id/eprint/37748/1/download%20%281%29.pdf>

96. Inclusion London. *Tackling Online Abuse: Petitions Commons Select Committee's Inquiry on Online Abuse and the Experience of Disabled People.* Published 2020. Accessed January 31, 2025.

<https://committees.parliament.uk/writtenevidence/9813/pdf/>

97. Johnson A. *How to Address Children's Online Safety in the United States.* Information Technology & Innovation Foundation. Published 2024. Accessed January 31, 2025.

<https://www2.itif.org/2024-child-online-safety.pdf>

98. Colliver C, King J. *The First 100 Days: Coronavirus and Crisis Management on Social Media Platforms.* ISD. Published 2020. Accessed January 10, 2025.

<https://www.isdglobal.org/wp-content/uploads/2020/06/First-100-Days.pdf>

99. Bundtzen S. *Misogynistic Pathways to Radicalisation: Recommended Measures to Assess and Mitigate Online Gender-Based Violence.* Institute of Strategic Dialogue. Published 2023.

Accessed January 31, 2025.

<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>

100. Institute for Strategic Dialogue [ISD]. *TikTok Series: Policy Recommendations.* Published 2024. Accessed January 31, 2025.

https://www.isdglobal.org/digital_dispatches/trouble-with-tiktok-policy-recommendations/

101. Gallagher A, Cooper L, Bhatnagar R, Gatewood C. *Pulling Back the Curtain: An Exploration*

of YouTube's Recommendation Algorithm. ISD. Published 2024. Accessed January 31, 2025. <https://www.isdglobal.org/wp-content/uploads/2024/06/Pulling-Back-the-Curtain-Executive-Summary-6.pdf>

102. Institute for Strategic Dialogue (ISD), B'nai B'rith International. *Online Antisemitism: A Toolkit for Civil Society*. UNESCO; Published 2022. Accessed January 10, 2025. <https://unesdoc.unesco.org/ark:/48223/pf0000381856>

103. The Jed Foundation [JED]. *The Jed Foundation (JED) Recommendations for Safeguarding Youth Well-Being on Social Media Platforms*. Published 2024. Accessed January 31, 2025. [https://jedfoundation.org/the-jed-foundation-jed-recommendations-for-safeguarding-youth-well-being-on-social-media-platforms/#:~:text=JED%20issues%20the%20following%20recommendations,alerts%20\(e.g.%2C%20push%20notifications\)](https://jedfoundation.org/the-jed-foundation-jed-recommendations-for-safeguarding-youth-well-being-on-social-media-platforms/#:~:text=JED%20issues%20the%20following%20recommendations,alerts%20(e.g.%2C%20push%20notifications))

104. Kids Online Health and Safety Task Force. *Online Health and Safety for Children and Youth: Best Practices for Families and Guidance for Industry*. Published 2024. Accessed January 31, 2025. <https://www.samhsa.gov/sites/default/files/online-health-safety-children-youth-report.pdf>

105. National Academies of Sciences, Engineering, and Medicine. *Social Media and Adolescent Health*. Washington, DC: The National Academies Press. Published 2024. Accessed January 31, 2025. <https://doi.org/10.17226/27396>

106. Sharrock S, Hudson N, Kerr J, Chalker C, David M, Myers C-A. *Key Attributes and Experiences of Cyberbullying Among Children in the UK*. National Centre for Social Research, Ofcom. Published 2023. Accessed January 31, 2025. <https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/keeping-children-safe-online/experiences-of-children/key-attributes-and-experiences-of-cyberbullying-among-children-in-the-uk.pdf?v=368017#:~:text=Repetition%20and%20intentionality%20were%20described,some%20participants%20as%20sometimes%20challenging>.

107. NYC Health. *New York City's Role in the National Crisis of Social Media and Youth Mental*

Health: Framework for Action. Published 2023. Accessed January 31, 2025.

<https://www.nyc.gov/assets/doh/downloads/pdf/mh/social-media-youth-mental-health-framework-action.pdf>

108. Office of the New York State Attorney General Letitia James. *Investigative Report on the Role of Online Platforms in the Tragic Mass Shooting in Buffalo on May 14, 2022*. Published 2022.

Accessed January 31, 2025. <https://ag.ny.gov/sites/default/files/buffaloshooting-onlineplatformsreport.pdf>

109. Stanley L, Tanner W, Treadwell J, Blagden J. *The Kids Aren't Alright: The 4 Factors Driving a Dangerous Detachment from Democracy*. UK Onward. Published 2022. Accessed January 31,

2025. <https://www.ukonward.com/reports/the-kids-arent-alright-democracy/>

110. Forland S, Meysenburg N, Solis E. *Age Verification: The Complicated Effort to Protect Youth Online*. Open Technology Institute. Published 2024. Accessed January 31, 2025.

<https://www.newamerica.org/oti/reports/age-verification-the-complicated-effort-to-protect-youth-online/>

111. Orygen Institute. *Responding to the Social and Economic Drivers of Youth Mental Health: Policy Lab*. Published 2024. Accessed January 10, 2025. <https://www.orygen.org.au/Orygen-Institute/Policy-Areas/Social-and-environmental-factors/2024-Policy-Lab-responding-social-economic-drivers.aspx?ext=.pdf>

[economic-drivers.aspx?ext=.pdf](https://www.orygen.org.au/Orygen-Institute/Policy-Areas/Social-and-environmental-factors/2024-Policy-Lab-responding-social-economic-drivers.aspx?ext=.pdf)

112. Orygen Institute & Headspace. *Social Media and Australian Society*. Published 2024.

Accessed January 31, 2025.

https://headspace.org.au/assets/Advocacy-and-policy-submissions/Orygen-headspace-submission-Joint-Select-Committee-on-Social-Media-and-Australian-Society_v2.pdf

113. House of Representatives Standing Committee on Social Policy and Legal Affairs. *Protecting the Age of Innocence*. Commonwealth of Australia. Published 2020. Accessed January 31, 2025.

https://www.aph.gov.au/Parliamentary_Business/Committees/House/Social_Policy_and_Legal_Affai

[rs/Onlineageverification/Report](#)

114. Economics References Committee. *Influence of International Digital Platforms*. Parliament of the Commonwealth of Australia; Published 2023. Accessed January 31, 2025.

https://parlinfo.aph.gov.au/parlInfo/download/committees/reportsen/RB000119/toc_pdf/

[Influenceofinternationaldigitalplatforms.pdf](#)

115. 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; Published 2024.

Accessed January 31, 2025. [https://nest.greenant.net/index.php/s/QiR56KZpQzMPPBn?](https://nest.greenant.net/index.php/s/QiR56KZpQzMPPBn?mc_cid=8b145254ee&mc_eid=67df8ff35d)

[mc_cid=8b145254ee&mc_eid=67df8ff35d](#)

116. Public Policy Forum. *Canadian Commission on Democratic Expression. How to Make Online Platforms More Transparent and Accountable to Canadian Users*. Democratic Expression & Public Policy Forum; Published 2022. Accessed January 31, 2025.

[https://static1.squarespace.com/static/5ea874746663b45e14a384a4/t/](https://static1.squarespace.com/static/5ea874746663b45e14a384a4/t/62794492003e16063fab7552/1652114580390/DemX+2+-+English+-+May+4.pdf)

[62794492003e16063fab7552/1652114580390/DemX+2+-+English+-+May+4.pdf](#)

117. Cox K, Ogden T, Jordan V, Paille P. *COVID-19, Disinformation and Hateful Extremism*. RAND Europe; Published 2021. Accessed January 31, 2025. [https://assets.publishing.service.gov.uk/](https://assets.publishing.service.gov.uk/media/60c868448fa8f57ce58ce901/)

[media/60c868448fa8f57ce58ce901/](#)

[RAND Europe Final Report Hateful Extremism During COVID-19 Final.pdf](#)

118. Chaudhary C. *Harnessing the Feed: Social Media for Mental Health Information and Support*. ReachOut; Published 2024. Accessed January 31, 2025.

<https://d1robvhmkdpun.cloudfront.net/46e49639306f108e035212644ba15d45.pdf>

119. Williams D, McIntosh A, Farthing R. *Profiling Children for Advertising: Facebook's Monetisation of Young People's Personal Data*. Reset Tech Australia; Published 2021. Accessed January 31, 2025.

https://au.reset.tech/uploads/resettechaustralia_profiling-children-for-advertising-1.pdf

120. Reset Australia. *The Future of Digital Regulation in Australia: Five Policy Principles for a Safer Digital World*. Published 2022. Accessed January 31, 2025. <https://au.reset.tech/uploads/the-future-of-digital-regulations-in-australia.pdf>
121. Reset Australia. *Achieving Digital Platform Public Transparency in Australia*. Published 2024. Accessed January 31, 2025. <https://au.reset.tech/Digital-platform-public-transparency.pdf>
122. Reset Australia. *Best Interests & Targeting: Implementing the Privacy Act Review to Advance Children's Rights*. Published 2024. Accessed January 31, 2025. <https://au.reset.tech/uploads/Best-Interests-Report-240128-digital.pdf>
123. Reset Australia. *Digital Platform Regulation Green Paper*. Published 2024. Accessed January 31, 2025. <https://au.reset.tech/uploads/Digital-Platform-Regulation-Green-Paper.pdf>
124. 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. Published 2022. Accessed January 31, 2025. <https://au.reset.tech/uploads/insta-pro-eating-disorder-bubble-april-22-1.pdf>
125. Dubicka B, Theodosiou L. *Technology Use and the Mental Health of Children and Young People*. Royal College of Psychiatrists. Published 2020. Accessed January 31, 2025. <https://www.rcpsych.ac.uk/docs/default-source/improving-care/better-mh-policy/college-reports/college-report-cr225.pdf>
126. Singh A, Jooshandeh J. *Platforms and the Public Square: A Taxonomy of Misinformation and the Misinformed*. Royal Society for Arts & BT Group. Published 2021. Accessed January 31, 2025. https://www.thersa.org/globalassets/foundation/new-site-blocks-and-images/reports/2021/11/platforms_and_the_public_square.pdf
127. Samaritans. *Towards a Suicide-Safer Internet*. Published 2022. Accessed January 31, 2025. https://nspa.org.uk/wp-content/uploads/2022/02/Samaritans_WhatASafeInternetLooksLike_2022.pdf
128. The Centre for Social Justice. *Unsafe Children: Driving Up Our Country's Response to Child*

Sexual Abuse and Exploitation. Published 2021. Accessed January 31, 2025.

<https://www.centreforsocialjustice.org.uk/wp-content/uploads/2021/03/CSJJ8804-Unsafe-Children-210325-WEB.pdf>

129. The Office of Minnesota Attorney General Keith Ellison. *Minnesota Attorney General's Report on Emerging Technology and Its Effects on Youth Well-Being*. Published 2024. Accessed January 31, 2025. <https://www.lrl.mn.gov/docs/2024/mandated/240175.pdf>

130. UNICEF Australia. *Protecting Children in the Online World - Reshaping the Digital World for Aussie Kids*. Published 2024. Accessed January 31, 2025.

131. The U.S. Surgeon General's Advisory. *Protecting Youth Mental Health*. Published 2021. Accessed January 31, 2025. <https://www.hhs.gov/sites/default/files/surgeon-general-youth-mental-health-advisory.pdf>

132. The U.S. Surgeon General's Advisory. *Our Epidemic of Loneliness and Isolation*. Washington (DC): US Department of Health and Human Services. Published 2023. Accessed January 31, 2025. <https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf>

133. WHO. *Teens, Screens and Mental Health*. Published 2024. Accessed January 31, 2025. <https://www.who.int/europe/news-room/25-09-2024-teens--screens-and-mental-health>

134. Livingstone S, Stoilova M, Stănicke LI, et al. *Young People Experiencing Internet-Related Mental Health Difficulties: The Benefits and Risks of Digital Skills*. ySkills. Published 2022. Accessed January 31, 2025. doi:10.5281/zenodo.6976423

135. West D. How disinformation defined the 2024 election narrative. *Brookings*. Published 2024. Accessed January 31, 2025. <https://www.brookings.edu/articles/how-disinformation-defined-the-2024-election-narrative/>

136. Duféal AL. *Norway Aims to Raise Age Limit for Social Media Users*. Published 2024. Accessed January 31, 2025. <https://brusselssignal.eu/2024/10/norway-aims-to-raise-age-limit-for-social-media-users/>

137. Walsh A. TikTok Stops Working as US Ban Comes Into Force. *BBC*. Published January 19 2025. Accessed January 31, 2025. <https://www.bbc.com/news/articles/cz6p1g54q85o>
138. Gillett R, Stardust Z, Burgess J. *Safety for Whom? Investigating How Platforms Frame and Perform Safety and Harm Interventions*. *Social Media+ Society*. 2022;8(4):20563051221144315.
139. Paul K. Reversal of Content Policies at Alphabet, Meta, and X Threaten Democracy, Warn Experts. *The Guardian*. Published 2023. Accessed January 31, 2025. <https://www.theguardian.com/media/2023/dec/07/2024-elections-social-media-content-safety-policies-moderation>
140. Latiff R. *ByteDance's TikTok Cuts Hundreds of Jobs in Shift Towards AI Content Moderation*. *Reuters*. Published 2024. Accessed January 31, 2025. <https://www.reuters.com/technology/bytedance-cuts-over-700-jobs-malaysia-shift-towards-ai-moderation-sources-say-2024-10-11/>
141. Fried I. Meta's New Policies Open Gate to Hate. *Axios*. Published January 9 2025. Accessed January 31, 2025. <https://www.axios.com/2025/01/09/meta-moderation-transgender-women-hate>
142. Gilbert D. Meta ditches fact-checkers ahead of Trump's second term. *Wired*. Published Jan 7 2025. Accessed January 31, 2025. <https://www.wired.com/story/meta-ditches-fact-checkers-in-favor-of-x-style-community-notes/>
143. Draper D, Neschke S. *The pros and cons of social media algorithms*. Bipartisan Policy Center. Published October 2023. Accessed January 31, 2025. https://bipartisanpolicy.org/download/?file=/wp-content/uploads/2023/10/BPC_Tech-Algorithm-Tradeoffs_R01.pdf
144. Ramsay Mental Health. Addicted to bad news: How 'doomscrolling' is affecting your mental health. Published 2021. Accessed January 31, 2025. <https://www.ramsaymentalhealth.com.au/en/resources-support/addicted-to-bad-news/>
145. University Hospitals. Doomscrolling: Breaking the habit. Published July 9, 2024. Accessed January 31, 2025. <https://www.uhhospitals.org/blog/articles/2024/07/doomscrolling-breaking-the->

[habit#:~:text=The%20effects%20of%20doomscrolling%20on,can%20also%20trigger%20panic%20attacks](#)

146. eSafety Commissioner. *Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography*. Published August 2023. Accessed January 31, 2025. <https://www.esafety.gov.au/sites/default/files/2023-08/Age-verification-background-report.pdf?v=1736394644828>

147. Huddleston J. *Improving youth online safety without sacrificing privacy and speech*. CATO Institute. Published 2024. Accessed January 31, 2025. <https://www.cato.org/briefing-paper/improving-youth-online-safety-without-sacrificing-privacy-speech>

148. Taylor J. Social media age restrictions may push children online in secret, Australian eSafety commissioner says. *The Guardian*. Published June 23, 2024. Accessed January 31, 2025. <https://www.theguardian.com/australia-news/article/2024/jun/23/social-media-age-restrictions-may-push-children-online-in-secret-australia-regulator-says>

149. Karp P, Touma R. Social media platforms with 'low risk of harm to children' could escape Albanese government age ban. *The Guardian*. Published October 10, 2024. Accessed January 31, 2025. <https://www.theguardian.com/australia-news/2024/oct/10/australia-social-media-ban-companies-low-harm-risk-michelle-rowland>

150. Giardini T, Buza M. Global digital policy roundup: November 2024. *Tech Policy Press*. Published November 2024. Accessed January 31, 2025. <https://www.techpolicy.press/global-digital-policy-roundup-november-2024/>

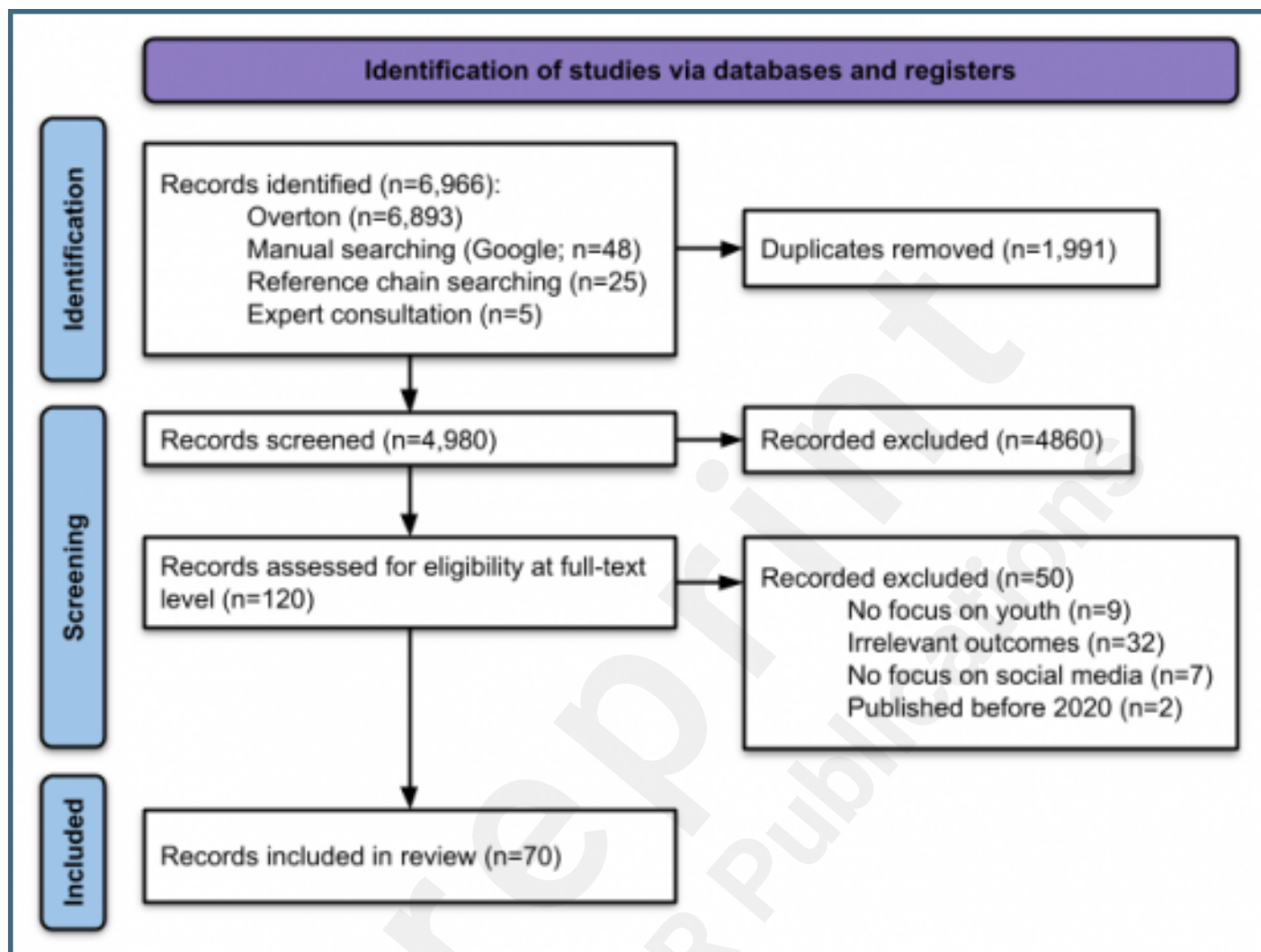
151. Morten CJ, Nicholas G, Viljoen S. Researcher access to social media data: Lessons from clinical trial data sharing. *Berkeley Tech Law J*. 2024;39(1):109-204. [doi:10.15779/Z38QF8JK81](https://doi.org/10.15779/Z38QF8JK81)



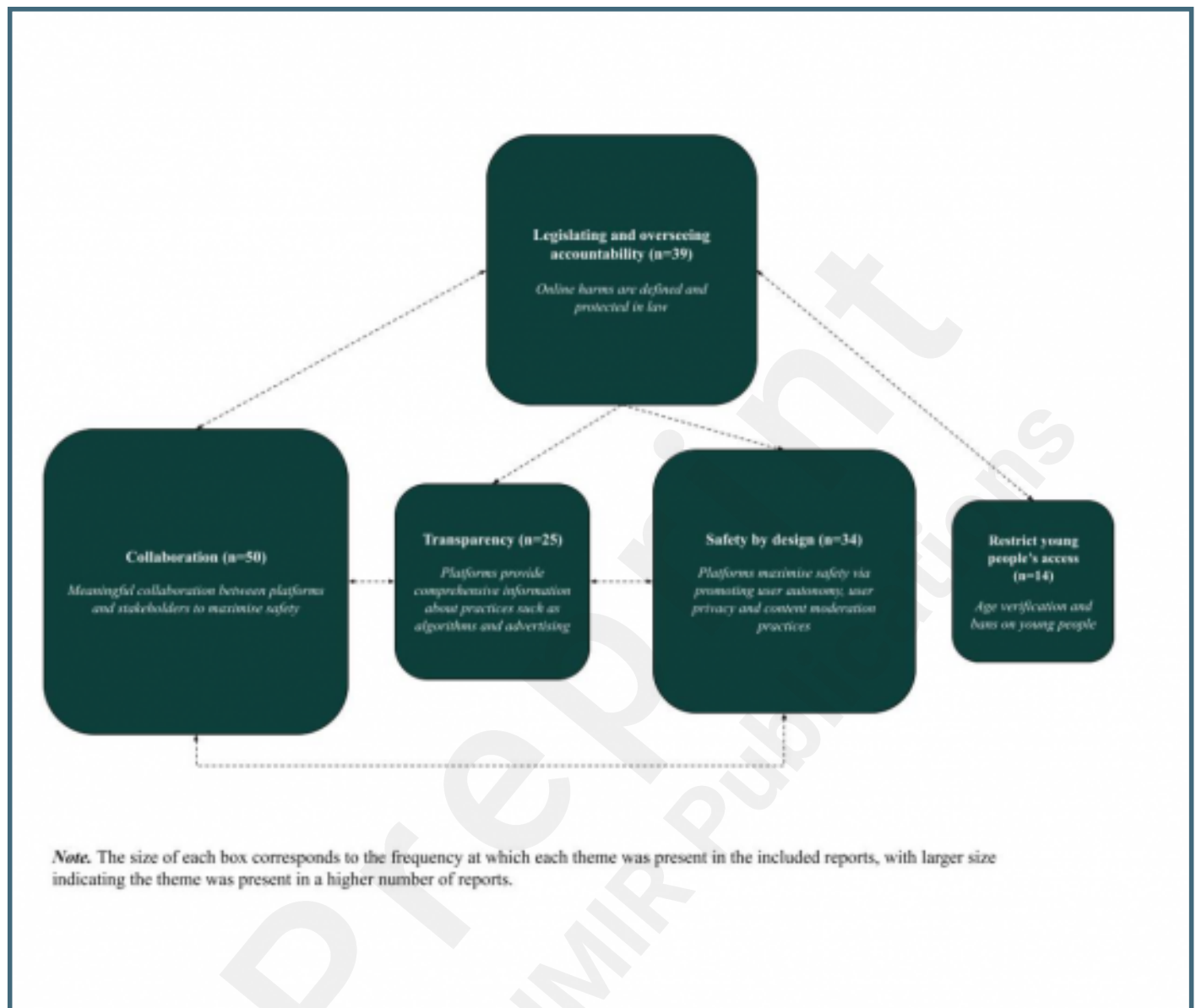
Supplementary Files

Figures

PRISMA Diagram.



Thematic map of the five core interrelated themes of social media recommendations.



CONSORT (or other) checklists

PRISMA-ScR Checklist.

URL: <http://asset.jmir.pub/assets/4feca5efad23936c61d630436b6106ea.pdf>

