



Kent Academic Repository

Mbioh, Will (2026) *Social media age-gating: does it reduce harm? UK Online Safety Act, EU Digital Services Act and Australia's social media ban*. Information & Communications Technology Law . pp. 1-13. ISSN 1360-0834.

Downloaded from

<https://kar.kent.ac.uk/116095/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1080/13600834.2026.2721015>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY-NC-ND (Attribution-NonCommercial-NoDerivatives)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).



Social media age-gating: does it reduce harm? UK Online Safety Act, EU Digital Services Act and Australia's social media ban

Will Mbioh

To cite this article: Will Mbioh (12 Sep 2026): Social media age-gating: does it reduce harm? UK Online Safety Act, EU Digital Services Act and Australia's social media ban, Information & Communications Technology Law, DOI: [10.1080/13600834.2026.2721015](https://doi.org/10.1080/13600834.2026.2721015)

To link to this article: <https://doi.org/10.1080/13600834.2026.2721015>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 12 Sep 2026.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Social media age-gating: does it reduce harm? UK Online Safety Act, EU Digital Services Act and Australia's social media ban

Will Mbioh

Kent Law School, University of Kent, Canterbury, UK

ABSTRACT

This paper evaluates whether social-media age-gating – Australia's under-16 ban – can reduce harm relative to the UK Online Safety Act and the EU Digital Services Act. Bringing clinical, developmental, and behavioural evidence into a legal analysis, it shows that risk is heavy-tailed and co-produced: a minority with pre-existing vulnerabilities experience severe outcomes amplified by platform affordances (visibility, metrics, ranking, availability). Because 'time on platform' is endogenous – simultaneously predictor, mediator, and outcome – simple dose-response claims fail, and exclusion by age mostly postpones first contact while leaving the state-behaviour-design loop intact. The paper argues that age-gating alone is unlikely to change whether severe cases occur; it changes when they appear. It proposes a two-pillar response: (1) investment in adolescent mental-health capacity (school-embedded early intervention, anti-bullying, family support, offline alternatives, algorithmic literacy) and (2) targeted design duties (default night-mode, diversity injections in sensitive niches, audited tail-focused studies), given legal force based on the DSA as the template.

KEYWORDS

Social media age-gating; Online Safety Act (OSA); Digital Services Act (DSA); youth online safety; recommender systems; adolescent mental health; heavy-tailed harms

Introduction

Over the past decade, policy on youth online safety has moved from simply content-gating (i.e. restricting access to categories such as pornography, extreme violence, or self-harm material – the core logic of the UK's Online Safety Act and, with additional system-design duties, much of the EU's Digital Services Act) to asking whether minors should be on major social-media and video-sharing platforms at all; that is, service-gating, or, barring account creation or participation below a set age – a newer, more radical approach exemplified by Australia's social-media minimum-age reforms.¹ Service-gating, in particular, promises a simple answer to a complex problem: if

CONTACT Will Mbioh  W.R.Mbioh@kent.ac.uk

¹Y Jang and B Ko, 'Online Safety for Children and Youth under the 4Cs Framework – A Focus on Digital Policies in Australia, Canada, and the UK' (2023) 10 Children 1415; Organisation for Economic Co-operation and Development, 'The Legal and Policy Landscape of Age Assurance Online for Child Safety and Well-Being' (OECD Publishing, 2025) <<https://>

© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

participation is postponed, the harms associated with highly visual, interactive platforms (e.g. mood and anxiety problems; intensified social comparison and body dissatisfaction; compulsive checking and sleep disruption; harassment, loneliness and family conflict; feeds that normalise self-harm or disordered-eating routines)² should recede.

This paper tests that premise. It asks whether exclusion by age is consistent with the mechanisms of harm identified in clinical, developmental and behavioural evidence, or whether it mainly defers first contact while leaving those mechanisms, and the underlying susceptibilities, intact. That is, rather than revisiting the debates that have dominated discussion since Australia's proposal – i.e. circumvention by users, platform designation and boundary-drawing, and other compliance-classification issues³ – the paper poses a prior question: does an age threshold map onto what the best available clinical, developmental and behavioural evidence says about who is harmed, how harm arises, and when it escalates? The analysis foregrounds that evidence and evaluates service-gating against it.

It contends that the evidence points to a heavy-tailed, co-produced, bidirectional problem: heavy-tailed in the sense that most teenagers report mixed or modest effects;⁴ a concentrated minority experience severe, impairing outcomes.⁵ These arise

www.oecd.org/en/publications/the-legal-and-policy-landscape-of-age-assurance-online-for-child-safety-and-well-being_4a1878aa-en.html accessed 8 October 2025.

²Maria T Maza and others, 'Association of Habitual Checking Behaviors on Social Media with Longitudinal Functional Brain Development' (2023) 177 JAMA pediatrics 160; S Choukas-Bradley and others, 'The Perfect Storm: A Developmental–Sociocultural Framework for the Role of Social Media in Adolescent Girls' Body Image Concerns and Mental Health' (2022) 25 Clinical Child and Family Psychology Review 681; Yvonne Kelly and others, 'Social Media Use and Adolescent Mental Health: Findings From the UK Millennium Cohort Study' (2018) 6 Eclinicalmedicine 59; A Marchant and others, 'A Systematic Review of the Relationship between Internet Use, Self-Harm and Suicidal Behaviour in Young People: The Good, the Bad and the Unknown' (2017) 12 PLoS ONE e0181722; 'Mental Health of Children and Young People in England, 2023 – Wave 4 Follow up to the 2017 Survey: Part 1 – Mental Health' (NHS England Digital 2023) <<https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2023-wave-4-follow-up/part-1-mental-health>> accessed 2 October 2025; David S Bickham, 'Current Research and Viewpoints on Internet Addiction in Adolescents' (2021) 9 Current pediatrics reports 1; Xiongfei Cao and others, 'Understanding the Mechanism of Social Media Addiction: A Socio-Technical Systems Perspective' [2023] Information Development 02666669231218685; Cecilia Cheng, Peizhen Sun and Kwok-Kei Mak, 'Internet Addiction and Psychosocial Maladjustment: Avoidant Coping and Coping Inflexibility as Psychological Mechanisms' (2015) 18 Cyberpsychology, Behavior, and Social Networking 539; Shilpa Suresh Bisen and Yogesh M Deshpande, 'Understanding Internet Addiction: A Comprehensive Review' (2018) 23 Mental Health Review Journal 165.

³Ange Lavoipierre, 'Legal Fights Are Brewing over Which Online Social Platforms Will Have to Ban Under-16s' ABC News (23 September 2025) <<https://www.abc.net.au/news/2025-09-24/digital-dilemma-social-media-age-ban-platforms/105807302>> accessed 8 October 2025; Tiffanie Turnbull, 'Enforcing Australia's Social Media Ban on Kids Is Possible but Contains Risks, Report Says' BBC News (1 September 2025) <<https://www.bbc.co.uk/news/articles/clyrkj0d44vo>> accessed 8 October 2025; Josh Taylor, 'Australia's under 16s Social Media Ban Could Extend to Reddit, Twitch, Roblox and Even Dating Apps' The Guardian (24 September 2025) <<https://www.theguardian.com/media/2025/sep/24/australia-under-16-social-media-ban-could-extend-reddit-twitch-roblox-lego-play-steam-dating-apps#:~:text=Reddit%20and%20X%20are%20among,Instagram%2C%20TikTok%2C%20YouTube%20and%20X>> accessed 8 October 2025.

⁴Patti M Valkenburg, Adrian Meier and Ine Beyens, 'Social Media Use and Its Impact on Adolescent Mental Health: An Umbrella Review of the Evidence' (2022) 44 Current opinion in psychology 58.

⁵Jasmine Fardouly, 'Potential Effects of the Social Media Age Ban in Australia for Children Younger than 16 Years' (2025) 7 The Lancet Digital Health e235; Marchant and others (n 2); Elias Balt and others, 'Social Media Use of Adolescents Who Died by Suicide: Lessons from a Psychological Autopsy Study' (2023) 17 Child and adolescent psychiatry and mental health 48; Mohammed Saud Alsaïdan and others, 'The Prevalence and Determinants of Body Dysmorphic Disorder among Young Social Media Users: A Cross-Sectional Study' (2020) 12 Dermatology reports <<https://doi.org/10.4081/dr.2020.8685>> accessed 28 March 2022; Monica Gupta, Amita Jassi and Georgina Krebs, 'The Association between Social Media Use and Body Dysmorphic Symptoms in Young People' (2023) 14 Frontiers in Psychology 1231801; Grace Haddad, 'The Impact of Social Media on Body Dissatisfaction and Dysmorphia in Adolescents in the US' (PhD Thesis, University of Pittsburgh 2022); C Anderson and others, 'Bidirectional Relationships between Muscularity-Oriented Disordered Eating and Mental Health Constructs: A Prospective Study' [2024] Eating Disorders <<https://doi.org/10.1080/10640266.2024.2319947>>; P Aparicio-Martinez and others, 'Social Media, Thin-Ideal, Body Dissatisfaction and Disordered Eating Attitudes: An Exploratory Analysis' (2019) 16 International Journal of Environmental Research

from interactions between person, content, timing and design mediated by household buffers and wider structures.⁶ Adolescents bring a characteristic neurodevelopmental profile (heightened reward reactivity, still-maturing control),⁷ a high-stakes social ecology, and often disrupted sleep.⁸ Platforms supply visibility, metrics, ranking and availability that tune to those sensitivities.⁹ ‘Time on platform’ is therefore endogenous: distress increases minutes, minutes reshape feeds and sleep, and both loop back into distress.¹⁰

Accordingly, the paper argues that harms are co-produced by several interacting elements: who the young person is and how they are doing; what the feed serves and how closely it resonates with their concerns; when use happens and in what immediate circumstances; and how the system is built – its features, ranking, and public metrics – while all of this is further shaped by buffers, such as of family and school supports, and by wider social conditions such as inequality and mental health service capacity. Effective governance, therefore, must adjust these variables together. Age-gating can buy time, but without parallel design duties and sustained investment in child and youth mental health, it is more likely to change when severe cases emerge than whether they do.

In terms of structure, the paper is organised into three parts. Part I maps the governance typology and situates age-gating within existing regimes: the UK’s Online Safety Act, the EU’s Digital Services Act, and Australia’s minimum-age reforms. It identifies thresholds, duties, and enforcement mechanisms, and frames the testable premise against which service-gating is assessed. Part II synthesises clinical, developmental, and behavioural evidence on the distribution and mechanisms of harm – heavy-tailed outcomes, bidirectional state–behaviour–design loops, adolescent neurodevelopment and social context, and the endogeneity of ‘time on platform’ – and develops a co-production lens, using a wildfire analogy to clarify precipitant, accelerant, and maintainer roles. Part III translates mechanisms into policy: a two-pillar package combining investment in child and youth mental-health capacity (school-embedded early intervention, anti-bullying, family supports, non-screen activities, algorithmic literacy) with targeted design duties (default night/‘bedtime’ mode, diversity injections in sensitive niches, audited heavy-tail studies), and

and Public Health 4177; SE Mazzeo and others, ‘Mitigating Harms of Social Media for Adolescent Body Image and Eating Disorders: A Review’ [2024] *Psychology Research and Behavior Management*; Yitshak Alfasi, ‘The Grass Is Always Greener on My Friends’ Profiles: The Effect of Facebook Social Comparison on State Self-Esteem and Depression’ (2019) 147 *Personality and Individual Differences* 111; Krishna C Bathina and others, ‘Individuals with Depression Express More Distorted Thinking on Social Media’ (2021) 5 *Nature Human Behaviour* 458; C McAllister and others, ‘Associations between Adolescent Depression and Self-Harm Behaviors and Screen Media Use in a Nationally Representative Time-Diary Study’ (2021) 49 *Research on Child and Adolescent Psychopathology* 1623.

⁶Fardouly (n 5); Vladimir L Malygin and others, ‘Bio-Psycho-Social Model of Internet Addiction Behavior in Adolescents. Risk Factors and Principles of Therapy’ (2017) 49 *Neurology Bulletin* 88.

⁷Malygin and others (n 6); Olivia Best and Sasha Ban, ‘Adolescence: Physical Changes and Neurological Development’ (2021) 30 *British Journal of Nursing* 272; Michaela Åström, ‘The Adolescent Brain on Social-Media: A Systematic Review’; Ellen Barry, ‘Social Media Use Is Linked to Brain Changes in Teens, Research Finds’ *The New York Times* (3 January 2023) <<https://www.nytimes.com/2023/01/03/health/social-media-brain-adolescents.html>> accessed 18 February 2023.

⁸Meyran Boniel-Nissim and others, ‘Adolescent Use of Social Media and Associations with Sleep Patterns across 18 European and North American Countries’ (2023) 9 *Sleep Health* 314; Louise AS Brautsch and others, ‘Digital Media Use and Sleep in Late Adolescence and Young Adulthood: A Systematic Review’ (2023) 68 *Sleep Medicine Reviews* 101742; Rakesh Kumar Swain and Atanu Kumar Pati, ‘Use of Social Networking Sites (SNSs) and Its Repercussions on Sleep Quality, Psychosocial Behavior, Academic Performance and Circadian Rhythm of Humans—a Brief Review’ (2021) 52 *Biological Rhythm Research* 1139.

⁹Anderson and others (n 5); Kelly and others (n 2); Fardouly (n 5).

¹⁰Anderson and others (n 5).

shows how to give these force in law (DSA Arts 27, 34–35, 40; targeted amendments to the UK OSA; complementary duties in Australia). The conclusion returns to the policy test: shrink the pool, damp the loops, thicken the buffers.

Part I: typology of online-safety governance and the place of age-gating

Across comparable democracies, three dominant modes of online-safety governance in relation to minors have emerged.¹¹ The first, content-gating, derives from broadcast-era principles: use is permitted, but children's exposure to specified categories of content is restricted through duties on providers to identify, assess, and deploy proportionate measures to mitigate risk (i.e. children may access the service, but exposure is 'gated'). In the United Kingdom, the Online Safety Act 2023 (OSA) operationalises this via children's access assessments and children's risk assessments for in-scope services, with corresponding mitigation duties where children are likely to access the service (e.g. OSA s.36 (children's access assessments) and s.37 (children's risk assessments), read with Part 5 generally).¹² Providers must carry out these assessments to determine whether their service is likely to be accessed by children and, if so, deploy proportionate measures to reduce the risk that children encounter content that is harmful to them (Part 5 duties, including the obligation to operate proportionate systems and processes). The Act distinguishes between 'primary priority content' harmful to children (i.e. a highest-risk subset specified in the Act and secondary legislation, which services must prevent children from encountering) and 'priority content' harmful to children (a broader set requiring mitigation rather than absolute prevention) (see OSA ss.60–62: s.60 (content that is harmful to children), s.61 (primary-priority content), s.62 (priority content)). Enforcement sits with Ofcom, the UK's independent communications regulator, which issues binding codes and guidance and can impose sanctions. A complementary (i.e. in addition to, not a substitute for) illegal-content regime applies to all users, regardless of age, through the Act's separate duties in relation to illegality.¹³

The European Union follows a broadly similar content-and-safety logic but adds a second mode, system-design risk mitigation – call it system-gating in this context, meaning use is permitted, but the systems themselves must be made safer by design.¹⁴ Under the Digital Services Act (DSA), Very Large Online Platforms/Very Large Online Search Engines must conduct systemic-risk assessments and implement risk-reduction measures (Arts 34–35) with specific attention to minors' well-being. The DSA also introduces targeted advertising restrictions: platforms may not present ads based on profiling using minors' personal data (Art 28). Platforms must provide transparency and meaningful user choice around recommender systems (Art 27) – that is, disclose

¹¹Jang and Ko (n 1); Organisation for Economic Co-operation and Development (n 1).

¹²Benjamin Farrand, 'How Do We Understand Online Harms? The Impact of Conceptual Divides on Regulatory Divergence between the Online Safety Act and Digital Services Act' [2024] *Journal of Media Law* 1; Clare McGlynn, Lorna Woods and Alexandros Antoniou, 'Pornography, the Online Safety Act 2023 and the Need for Further Reform' [2024] *Journal of Media Law* 1.

¹³Farrand (n 12); McGlynn, Woods and Antoniou (n 12).

¹⁴Farrand (n 12); Peter Church and Ceyhun Necati Pehlivan, 'The Digital Services Act (DSA): A New Era for Online Harms and Intermediary Liability' (2023) 4 *Global Privacy Law Review*.

the main parameters and offer at least one option not based on profiling – and adopt additional constraints on manipulative interface design (Art 25); i.e. prohibitions on deceptive or coercive ‘dark patterns’ that steer choices, with special regard to minors.¹⁵

Australia, on the other hand, has moved to service-gating by age (use itself is prohibited for under-16s). The Online Safety Amendment (Social Media Minimum Age) Act 2024 (Cth) amends the Online Safety Act 2021 to create a minimum participation age.¹⁶ New provisions define ‘designated social media services’ and ‘restricted access social media account,’ and impose a ‘reasonable steps’ duty on designated services to prevent people under 16 from having or accessing accounts. Related divisions provide for parent-authorised restricted accounts for 16–17-year-olds and empower the eSafety Commissioner to issue guidance and enforce compliance (e.g. the new ‘reasonable steps’ obligation and associated monitoring/enforcement machinery inserted by Sch 1).¹⁷

Thus, Australia’s age-ban approach is the sharpest departure: it assumes that deferring service participation to 16 will materially reduce serious harm, relative to UK/EU strategies that keep minors on services but constrain content and system design. Rather than re-litigating feasibility, privacy or proportionality at this stage (issues that dominate current regulatory discourse),¹⁸ the paper will test that premise against the mechanisms of harm: that is, does current evidence suggest that excluding under-16s will reduce incidence, pace or duration of severe cases once re-entry occurs? Or does it shift when first contact happens without altering the underlying engine?

Part II: the evidence on psychosocial harm: a heavy-tailed, bidirectional problem

Broadly, reported psychosocial harms associated with social media and video-sharing platform use cluster across five domains; the evidence is the strongest for highly visual, interactive, comparison-oriented services driven by recommender engines, meaning platforms that foreground images and short video, solicit posting, liking and commenting, display public metrics such as likes and follower counts, and personalise feeds.¹⁹ The first domain are affective harms: higher rates of low mood, anxiety, irritability and stress reactivity, often linked to cyberbullying, social evaluation and ‘FOMO.’²⁰ The second domain concerns cognitive harms: intensified social comparison, body dissatisfaction and negative self-schema (e.g. ‘appearance equals worth’), plus rumination and attentional capture that crowd out reflective thinking.²¹ The third domain relate to behavioural injuries: compulsive checking, disrupted sleep and circadian rhythm, reduced physical activity, academic disengagement, and, for some, symptom-congruent behaviours

¹⁵Farrand (n 12).

¹⁶Terry Flew and Derek Wilding, ‘The Turn to Regulation in Digital Communication: The ACCC’s Digital Platforms Inquiry and Australian Media Policy’ (2021) 43 Media, Culture & Society 48; Fardouly (n 5).

¹⁷B Nansen and others, ‘Children and Digital Wellbeing in Australia: Online Regulation, Conduct and Competence’ (2012) 6 Journal of Children and Media 237; Jang and Ko (n 1).

¹⁸Lavoipierre (n 3); Turnbull (n 3); Taylor (n 3).

¹⁹Valkenburg, Meier and Beyens (n 4); Davide Marengo and others, ‘Highly-Visual Social Media and Internalizing Symptoms in Adolescence: The Mediating Role of Body Image Concerns’ (2018) 82 Computers in Human Behavior 63; Choukas-Bradley and others (n 2); Cao and others (n 2).

²⁰Kelly and others (n 2); Marchant and others (n 2); McAllister and others (n 5); ‘Mental Health of Children and Young People in England, 2023 – Wave 4 Follow up to the 2017 Survey: Part 1 – Mental Health’ (n 2) 4.

²¹Choukas-Bradley and others (n 2); Alfasi (n 5); Marengo and others (n 19).

such as restrictive eating, purging, over-exercise or self-injury.²² The fourth concerns social/relational risks: exposure to harassment and exclusion, pressure to perform for public metrics, shifts from offline to online contact that can increase loneliness, and family conflict over limits and use.²³ Finally, we have clinical escalators: algorithmically reinforced feeds that surface self-harm or disordered-eating themes, normalise risky routines, and slow recovery by rewarding symptom-aligned posting and late-night engagement.²⁴

(a) Heavy-Tailed Outcomes and Bidirectional Dynamics

These harms are heterogeneous and often heavy-tailed: most teenagers experience mixed or modest effects (small day-to-day upsides and downsides without lasting functional impairment),²⁵ while a minority with pre-existing vulnerabilities (mood dysregulation, eating- and body-image concerns) face severe, impairing outcomes.²⁶ In heavy-tailed distributions, extreme outcomes, such as sustained depressive episodes, self-harm, or clinically significant eating-disorder trajectories, occur more often than a bell curve would predict, so cohort-wide averages (mean associations and pooled effect sizes) look small even as a disproportionate share of the overall burden – persistent impairment, referrals to Child and Adolescent Mental Health Services, emergency presentations, and intensive service use – sits in the tail, concentrated among a relatively small subset of teenagers at the far end of the risk distribution.²⁷

The evidence is also clear that risk is bidirectional: underlying distress (anxiety, low mood, dysphoria, disordered-eating or body-image preoccupations, insomnia) shapes what young people seek and how they engage, and those signals teach the system to deliver more of the same.²⁸ In that way, a young person's initial state does more than 'filter' experience; it sets the gain and operates as an attractor, drawing in cues that feel especially relevant.²⁹ Their clicks, watch time, and follows are then interpreted by ranking systems as reasons to prioritise those niches, and the niches mirror, validate, and elaborate the pre-existing beliefs and symptoms (for example, perfectionistic or appearance-centred schemas), making them feel more accurate and necessary. That is to say, the altered feed shifts what they notice and feel: heightening the preoccupation and weakening brakes such as sleep and metacognition (the capacity to notice, monitor and evaluate one's own thoughts and urges, and to adjust behaviour), which then generates further engagement signals back to the system.³⁰ Over time, this state-behaviour-design loop can deepen preoccupations, erode metacognitive control and

²²Maza and others (n 2); Brautsch and others (n 8); Boniel-Nissim and others (n 8); Kelly and others (n 2).

²³Marchant and others (n 2); Aparicio-Martinez and others (n 5); Mazzeo and others (n 5); Cao and others (n 2); Brautsch and others (n 8).

²⁴Marchant and others (n 2); Aparicio-Martinez and others (n 5); Mazzeo and others (n 5); Cao and others (n 2); Brautsch and others (n 8).

²⁵Valkenburg, Meier and Beyens (n 4).

²⁶ibid; Marchant and others (n 2); Kelly and others (n 2); Choukas-Bradley and others (n 2); Malygin and others (n 6).

²⁷McAllister and others (n 5); Balt and others (n 5); Mazzeo and others (n 5); Aparicio-Martinez and others (n 5).

²⁸Brautsch and others (n 8); Kumar Swain and Pati (n 8); Fardouly (n 5); Valkenburg, Meier and Beyens (n 4).

²⁹Valkenburg, Meier and Beyens (n 4).

³⁰Alsaidan and others (n 5); Gupta, Jassi and Krebs (n 5); Francesca C Ryding and Daria J Kuss, 'The Use of Social Networking Sites, Body Image Dissatisfaction, and Body Dysmorphic Disorder: A Systematic Review of Psychological Research' (2020) 9 *Psychology of Popular Media* 412.

sleep, and extend episode duration,³¹ even as many peers experience mixed or beneficial impacts.³²

(b) Age-Linked Bidirectionality: Neurodevelopment and Social Context

Two features of teenagerhood make this loop unusually strong: neurodevelopmental vulnerability and a high-stakes social context.³³ On the neurological side, reward and salience systems (for example, dopamine-rich ventral-striatal and amygdala circuits) are highly reactive to novelty, status and social approval, while prefrontal control systems that support planning, impulse inhibition and metacognition are still maturing.³⁴ Pubertal changes also raise the perceptual salience of the body, and a natural circadian phase delay pushes sleep later, so late-night use more easily erodes self-control the next day.³⁵ On the psychosocial side, identity work, peer evaluation, rising academic pressure and early romantic dynamics make feedback, comparison and belonging exceptionally important.³⁶

Because image and video-centric platforms package these cues (e.g. public metrics, rapid novelty, variable-ratio rewards, appearance-focused visuals, always-on social contact) teenagers do not simply spend more time there by chance; the environment is thus tuned to their developmental sensitivities.³⁷ This produces a selection effect: teenagers who are most reactive to social feedback are most attracted to the features that provide it, which in turn increases exposure to the contingencies that shape mood, sleep and self-evaluation.³⁸ For those with pre-existing vulnerabilities (anxiety, depressive tendencies, eating- and body-image concerns), the fit is even tighter: reassurance-seeking, body checking and perfectionistic tracking map neatly onto creator niches and platform tools, so heavier use is both more likely and more consequential.³⁹

Age therefore adds a second layer of bidirectionality: the adolescent profile increases the probability of heavy, socially reinforced engagement; heavy, feedback-driven engagement amplifies the very processes – reward reactivity, sleep loss, rumination – that weaken prefrontal brakes and heighten distress.⁴⁰ The result is a concentrated risk pattern in which the users most likely to be adversely affected are also those most likely to be captured and kept by the system, not primarily because of ‘screen time’ per se, but because of a deep fit between developmental neurobiology, psychosocial priorities and the core affordances these platforms optimise.⁴¹

(c) Time as an Endogenous Variable: Rethinking Causation

³¹ Alsaidan and others (n 5); Gupta, Jassi and Krebs (n 5); Ryding and Kuss (n 30).

³² Fardouly (n 5).

³³ Malygin and others (n 6); Best and Ban (n 7); Åström (n 7); Barry (n 7).

³⁴ Malygin and others (n 6); Best and Ban (n 7); Åström (n 7); Barry (n 7).

³⁵ Malygin and others (n 6); Best and Ban (n 7); Åström (n 7); Barry (n 7).

³⁶ Sibel Altıkulaç and others, ‘The Teenage Brain: Public Perceptions of Neurocognitive Development during Adolescence’ (2019) 31 *Journal of Cognitive Neuroscience* 339; Suparna Choudhury, Tony Charman and Sarah-Jayne Blakemore, ‘Development of the Teenage Brain’ (2008) 2 *Mind, Brain, and Education* 142.

³⁷ Malygin and others (n 6); Best and Ban (n 7); Åström (n 7); Barry (n 7).

³⁸ Valkenburg, Meier and Beyens (n 4).

³⁹ Fardouly (n 5); Marchant and others (n 2); Balt and others (n 5); Alsaidan and others (n 5); Gupta, Jassi and Krebs (n 5); Haddad (n 5); Anderson and others (n 5); Aparicio-Martinez and others (n 5); Mazzeo and others (n 5); Alfasi (n 5); Bathina and others (n 5); McAllister and others (n 5).

⁴⁰ Maza and others (n 2); Brautsch and others (n 8); Boniel-Nissim and others (n 8); Choukas-Bradley and others (n 2).

⁴¹ Best and Ban (n 7); Choukas-Bradley and others (n 2); Anderson and others (n 5); Fardouly (n 5).

This correlation between adolescent development and platform design makes the relationship between time spent online and causation exceptionally complex, challenging any simple interpretation of the evidence.⁴² While studies frequently report a correlation between higher screen time and poorer mental health outcomes, the complex, bidirectional dynamics at play make it difficult to determine the causal direction of this link.⁴³ In that sense, ‘time on platform’ may sound like a tidy quantity, much like a drug dose, but it is not given from outside; it is *produced* by the same forces that drive harm.⁴⁴ When a teenager feels anxious, lonely, or dysphoric, they often turn to feeds for distraction, reassurance, or late-night company. That distress increases their minutes online, so time is, in part, an *outcome* of their symptoms; in other words, those extra minutes are a behavioural read-out of the underlying problem (coping, reassurance-seeking, avoidance), not a neutral exposure imposed on an otherwise steady state. Once online, the mix of comparison-heavy content, public metrics, and night-time scrolling can heighten arousal, undermine sleep, and sharpen preoccupations, which in turn worsens mood or body image.⁴⁵ In that role, time is a *mediator*; the pathway through which vulnerability affects later distress. And statistically, heavier use also *predicts* worse scores on wellbeing scales, so time appears as a *predictor*.⁴⁶ Because it simultaneously plays all three roles (i.e. predictor, mediator, and outcome) simple dose–response stories (‘more hours lead to more harm’) are unstable. Part of the ‘extra time’ we measure is actually a *signal of how unwell the young person already is* (e.g. compulsive checking during a low spell), not a clean cause of becoming unwell.⁴⁷ Without separating why the time increased, what was consumed, and when it occurred, correlations between hours and harm can reflect pre-existing distress as much as any causal effect of use.⁴⁸

This is another way to say that, in effect, time on platform is an *endogenous* variable: distress, social stakes, and algorithmic tightening increase use, while increased use alters distress and the feed that follows.⁴⁹ This means that time is determined *inside* the very system we are trying to study. A teenager’s distress, their current social pressures (falling out with friends, exams), and the algorithm’s tightening around what holds their attention all *push time upward*: feeling low leads to scrolling to cope; big peer drama results in checking more; highly tailored feed leads to its being harder to put down the phone.⁵⁰ But the arrow also runs the other way. *More time* changes the teenager’s state (sleep loss, more comparison, stronger urges) *and* teaches the recommender to supply denser, stickier content, which then alters tomorrow’s time again.⁵¹

Because minutes both result from risk factors and produce new risk, a simple ‘dose–response’ reading (‘more minutes cause more harm’) is misleading. Is heavy use harmful because of the extra minutes, or are extra minutes a signal that the teenager is already struggling and being algorithmically captured? Usually it’s both,

⁴²Fardouly (n 5).

⁴³Valkenburg, Meier and Beyens (n 4); Marchant and others (n 2); Anderson and others (n 5).

⁴⁴Cheng, Sun and Mak (n 2); Bickham (n 2); Cao and others (n 2).

⁴⁵Brautsch and others (n 8); Boniel-Nissim and others (n 8); Kelly and others (n 2); Choukas-Bradley and others (n 2); Marengo and others (n 19).

⁴⁶Kelly and others (n 2); McAllister and others (n 5); Valkenburg, Meier and Beyens (n 4).

⁴⁷Maza and others (n 2).

⁴⁸Kelly and others (n 2); McAllister and others (n 5); Valkenburg, Meier and Beyens (n 4).

⁴⁹Maza and others (n 2); Cheng, Sun and Mak (n 2); Bickham (n 2); Cao and others (n 2).

⁵⁰Cheng, Sun and Mak (n 2); Bickham (n 2); Choukas-Bradley and others (n 2).

⁵¹Boniel-Nissim and others (n 8); Brautsch and others (n 8); Marengo and others (n 19).

intertwined.⁵² One can imagine it as a loop: distress leads to more minutes, which means a tighter feed, which translates to more distress and to more minutes. Time is simultaneously *input and output*, making causal claims based on minutes alone unreliable.⁵³

Because ‘time’ is produced inside the loop, everything outside the loop that pushes on the teen or the feed also pushes on the causal story – and many of those forces are neither controllable nor constant.⁵⁴ Sleep debt, exam periods, puberty timing, bullying episodes, family stress, poverty, offline activity options, and school climate can all raise distress or change opportunities to cope; they simultaneously nudge engagement up and make the content that appears feel more compelling.⁵⁵ These are classic moving confounders and moderators: they both generate the conditions under which a teen uses a platform and change how that use lands. They also arrive as shocks (a breakup, a viral rumour), vary week to week, and spill over between people (peer drama, local trends), so we cannot ‘hold them fixed’ long enough to read a clean dose–response to platform exposure. Even platform-side tweaks interact with them: the same notification cadence is tolerable for a well-rested teen with strong offline supports but destabilising for a sleep-deprived teen during exams.⁵⁶ The upshot is co-production of outcomes – online design and offline context work together – so attributing harm solely to the platform, or solely to the background factors, will systematically misdescribe what is happening in real lives.⁵⁷

(d) Platforms as Co-Producers: Precipitant, Accelerant, Maintainer

Because of this, it is hard to see social media as a clean, isolatable ‘signal’ because, in the material context of users, it sits everywhere in the causal chain at once: it is the setting where things happen, the channel that carries them, and the tool people use to cope with what has happened. That makes it less like a single ingredient and more like the pan the whole meal is cooked in – shaping heat, timing, and flavour without being the food itself. Conceptually, it acts as a modulator or co-producer.⁵⁸

Framing platforms as co-producers/modulators of experience is more accurate and more useful than asking whether they ‘cause’ harm outright. Co-production means the platform’s affordances (visibility, metrics, ranking, availability) interact with a teen’s state and context to shape the tone, tempo, and trajectory of what they feel. Instead of flipping problems on or off, the system alters how intensely feelings register (salience), how quickly they escalate (pace), how long they linger (persistence), and what they seem to ‘mean’ (interpretation). Seen this way, for those already at risk at the tail, the platform can serve as a *precipitant*, tipping a loaded probability of risk into an episode by turning momentary distress and salient cues into tightly matched exposure at sensitive times. It can then act as an *accelerant* and, to a degree, a *maintainer* by shortening the interval between similar exposures, deepening engagement, and sustaining symptom-congruent routines over days and nights;

⁵²Kelly and others (n 2); McAllister and others (n 5); Valkenburg, Meier and Beyens (n 4).

⁵³Cao and others (n 2); Maza and others (n 2); Anderson and others (n 5).

⁵⁴Valkenburg, Meier and Beyens (n 4); Marchant and others (n 2); Kelly and others (n 2).

⁵⁵Malygin and others (n 6); Brautsch and others (n 8); Boniel-Nissim and others (n 8); Kelly and others (n 2); Choukas-Bradley and others (n 2).

⁵⁶Brautsch and others (n 8); Boniel-Nissim and others (n 8).

⁵⁷Valkenburg, Meier and Beyens (n 4); Cao and others (n 2); Cheng, Sun and Mak (n 2).

⁵⁸Valkenburg, Meier and Beyens (n 4); Marchant and others (n 2); Kelly and others (n 2); Cheng, Sun and Mak (n 2); Cao and others (n 2).

these are time-structured roles that interact with, and depend on, underlying traits and the surrounding social conditions rather than replace them.⁵⁹

By way of analogy, consider a dry forest on a windy evening. The dryness – weeks without rain, fine fuel on the ground – stands for a young person's underlying state and context (vulnerability and social pressures). A single spark from a match is the precipitant: on its own it might fizzle, but in these conditions it can catch quickly. The wind is the platform's affordances (e.g. visibility, metrics, ranking and availability) raising the flame's intensity and speed; that is the accelerant, shortening the time between new ignitions by blowing embers into fresh tinder. Once parts of the forest begin to smoulder underground in peat and roots, the fire persists even when surface flames look lower; this is the maintainer, analogous to symptom-congruent routines and late-night checking that keep the state going. Firebreaks, damp ground and rain are buffers: supportive adults, sleep, and timely help that slow or stop spread. The topography and weather are the wider structures – poverty, school climate, service capacity – that set how often dangerous configurations arise. In this picture the platform does not create the forest or the drought; it modulates salience (how hot), pace (how fast it runs), persistence (how long it smoulders), and interpretation (which way the wind pushes the fireline). The risk comes from their joint configuration over time.

Part III: what can be done to address this problem

Taken as a whole, this evidence implies that solving the psychosocial harms associated with social media use requires interventions that interrupt the feedback loops through which vulnerability, engagement, and ranking systems co-constitute each other, rather than policies that merely delay exposure.⁶⁰ Because 'time on platform' is produced inside the same loop that generates harm, measures that simply remove under-16s from services, service-gating, do not, if used alone, reconfigure the mechanism that makes time expand, nor the conditions that make certain content feel urgent and self-confirming once participation begins.⁶¹ Service-gating may reduce the number of early encounters and postpone first contact, but it leaves susceptibility states, social contingencies, and high-gain design intact; when access resumes, the learning systems and adolescent incentives that drive tight person–content matches re-engage.⁶² That is, the function (*the underlying state–behaviour–design loop that tightens person–content fit, expands use under distress, and teaches ranking systems to serve denser niches*) is unaltered even if the input, namely the first period of access and time that seed the recommender's learning, is delayed. In the terms of the earlier analogy, postponing participation lowers the chance of a spark today, but it does not thin the fuel, calm the wind, or carve firebreaks; when conditions return, ignition and spread remain likely.⁶³

A substantive solution, therefore, does not lie in simply holding back users, but in addressing the underlying sources of their vulnerability and re-engineering the digital

⁵⁹Fardouly (n 5).

⁶⁰Valkenburg, Meier and Beyens (n 4); Marchant and others (n 2); Kelly and others (n 2); Malygin and others (n 6).

⁶¹Maza and others (n 2); McAllister and others (n 5); Cheng, Sun and Mak (n 2); Boniel-Nissim and others (n 8); Brautsch and others (n 8).

⁶²Choukas-Bradley and others (n 2); Cao and others (n 2); Aparicio-Martinez and others (n 5); Mazzeo and others (n 5); Fardouly (n 5).

⁶³Jang and Ko (n 1); Organisation for Economic Co-operation and Development (n 1); Fardouly (n 5).

environment to be safer. This requires a two-pronged strategy that combines a foundational public-health intervention⁶⁴ with a targeted regulatory duty for platform design.⁶⁵

(a) Shrinking Susceptibility: Specialist Child and Adolescent Services, Anti-Bullying, and Offline Alternatives

The first and most essential pillar is a major investment in adolescent mental-health services. In practice, that means specialist child and adolescent services with a clear, single point of entry for referrals; brief screening within days and routine assessments within weeks; and multidisciplinary teams able to step care up or down without fragmented hand-offs. It also means schools with embedded early-intervention pathways: on-site teams delivering low-intensity support (guided self-help and brief cognitive-behavioural programmes for sleep, anxiety, and body-image concerns), open drop-ins, and rapid escalation routes into specialist or primary care when thresholds are met.

Families need structured support as well – practical psychoeducation and parent-led modules that help caregivers monitor sleep, set device routines, and recognise early warning signs, alongside clear navigation for what to do next. Moreover, school climate should be addressed through whole-school anti-bullying frameworks: a designated safeguarding/child-protection lead with named responsibility; simple, anonymous reporting tools; incident logs that are actually reviewed; periodic climate surveys; staff and bystander training; and proportionate responses that combine sanction with restorative options.

Outside the school day, adequate funding should keep non-screen recreation affordable and reachable – after-school sport, arts, and youth provision with fee waivers where needed and safe transport – so there are real alternatives to evening screen time. Finally, algorithmic literacy should be taught, not assumed: embedded within health/citizenship/computing curricula, with clear outcomes (how feeds rank and personalise; how public metrics shape behaviour; how to recognise ‘tightening’; how to set quiet hours and limit notifications; when and how to seek help), supported by teacher professional development, standardised lesson packs, and parent workshops, and assessed through scenario-based tasks rather than recall alone. In the analogy, these measures are the regular rainfall and damp ground that make sparks less likely to catch and easier to extinguish early.

(b) Dampening the Loops: Recommenders, Public Metrics, and Transparency Duties

The second pillar is to give legal force (drawing on the DSA’s systemic-risk approach) to a small number of specific, high-impact design and timing changes. The first is the provision of night/‘bedtime’ mode for under-18s (default and on by design). Technically, this means a server-side schedule keyed to the user’s local time so that, as evening progresses, the recommender’s refresh interval is lengthened (from frequent updates to progressively slower ones), background auto-refresh and prefetch are suspended, and non-essential push notifications (e.g. new likes, follows, suggested content, streak prompts) are first rate-limited and then paused. Public counters (likes, views, follower totals) are

⁶⁴Fardouly (n 5).

⁶⁵Farrand (n 12).

either hidden or held static overnight, so they do not update in real time. By a set cutoff (e.g. midnight), ranking updates stop for minor accounts: the feed does not reload in the background and new items do not auto-surface. If a user wants something after that point, they must issue an explicit search or follow a direct link. The point is to protect sleep and dampen late-night escalation without banning access.

The second is to introduce diversity injections in sensitive niches. Where topics are known risk zones (nutrition, body image, self-harm, mood), the ranking should reserve a portion of the feed for off-niche, non-confirmatory items and up-weight reputable, professional sources (for example, NHS and recognised clinical bodies), with clear source labels. Not by blocking searches, but ensuring that alongside creator content, users also see balanced, evidence-based material so a momentary preoccupation does not collapse the feed into a narrow, self-confirming tunnel or echo chamber. Practically, this can be implemented at the last step of ranking (the reordering stage) by enforcing a minimum mix rule for these topics and setting a cap on consecutive near-duplicate items in a session.

The third suggestion is audited well-being studies focused on the 'tail.' Large services should run independent, audited surveys and cohort studies of users under 18, with deliberate oversampling to capture those with pre-existing conditions (eating-disorder risk, mood dysregulation, sleep problems). Reports should include a small set of heavy-tail indicators – for example, late-night session length, the number of sessions after a set hour (e.g. 23:00), repeated engagement with symptom-congruent themes, counts of non-essential notifications delivered after 21:00, and the share of impressions from authoritative sources in sensitive niches – and should document which mitigations were triggered (night-mode active, notifications paused, diversity shown) when severe patterns appeared. Findings should be shareable with regulators and vetted researchers in privacy-preserving, aggregated form.

In terms of how to give this legal effect, the EU's DSA model could operate as the template. Articles 34–35 (systemic-risk assessment and mitigation) can be used to require the night-mode schedule and diversity injections in specified risk topics; Article 27 (recommender transparency) can be relied on to require disclosure of night-mode parameters (e.g. refresh intervals, cutoff times, notification categories paused) and diversity-injection logic; and Article 40 (data access) can support audited, tail-focused studies. For the UK, this would mean amending the Online Safety Act to make design-and-timing duties explicit within the children's safety regime in Part 5, empowering Ofcom to require (via binding codes) the night-mode schedule, sensitive-topic diversity injections with authoritative boosting, and audited tail-focused research. It would also mean expanding s.159 transparency so the largest services provide operational logs of night-mode (e.g. hours active, notifications suppressed) and diversity-injection performance (e.g. proportion of authoritative impressions), and share anonymised, researcher-grade data for independent evaluation. Likewise, Australia could retain the minimum-age rule but add complementary design-and-timing obligations to the Online Safety Act 2021 so the eSafety Commissioner can require the same night-mode schedule, diversity injections in sensitive topics, and audited heavy-tail studies with researcher access – so exclusion buys time and, on re-entry, the core mechanisms are safer.

Conclusion

To be clear, this paper is *not* arguing that service-gating should not be used. It argues that, on its own, it is a blunt instrument that defers first contact but leaves the mechanism that concentrates harm unchanged. Pairing any age rule with the specific design-and-timing controls above, plus serious investment in adolescent mental health and supports around school and home, is the evidence-consistent path to fewer ignitions, slower escalations, and shorter episodes for the minority most at risk, and a more sustainable load for families, schools, and clinical services.

Disclosure statement

No potential conflict of interest was reported by the author(s).