

Beyond Parental Control: Digital Literacy and Adolescent Online Safety in Malaysia

Mohamad Rizal Abd Rahman^{1*}, Wan Farahiyah Izni W Abd Rahman², Siti Farahiyah Ab Rahim³, Nur Farahiyah Mohd Nasir⁴, Masnizah Mohd⁵

¹Associate Professor, Faculty of Law, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.

²Senior Lecturer, Faculty of Law, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.

³Senior Lecturer, Faculty of Law, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.

⁴Senior Lecturer, Faculty of Law, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.

⁵Associate Professor, Centre for Cyber Security Studies, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia.

*Corresponding Author

DOI: <https://doi.org/10.47772/IJRISS.2026.100700612>

Received: 26 July 2026; Accepted: 01 August 2026; Published: 08 August 2026

ABSTRACT

Malaysia's transition into a highly connected digital society has intensified both the opportunities and risks experienced by adolescents. This article examines how parental mediation and digital literacy interact with online risk and evaluates whether a parent-centred model remains adequate following the commencement of the Online Safety Act 2025 and the introduction of social-media account restrictions for users below 16. Using a doctrinal and evidence-informed socio-legal approach, the study applies a transparent purposive selection framework to Malaysian legislation, subsidiary instruments, official statistics, published child-consultation evidence and peer-reviewed empirical research. Sources were included where they addressed adolescent participation, parental mediation, digital literacy, disclosure, platform design or institutional accountability, and were synthesised against four outcomes: exposure, competence, trust and disclosure, and access to remedy. The analysis finds that digital literacy has a dual effect: it expands learning and participation while also increasing encounters with risk. Restrictive mediation may reduce immediate exposure but can suppress skills, trust and disclosure, whereas active mediation and proportionate supervision are more compatible with resilience. Recent consultations involving children in Malaysia reinforce that online access is integral to education and connection, while young users may still feel unsafe, unheard and left to manage harm alone. The Online Safety Act 2025 shifts responsibility towards platforms through duties concerning safe design, reporting, assistance and child-user protection. The article develops a Shared Digital Responsibility Model that assigns distinct obligations to adolescents, caregivers, schools, platforms and regulators, and operationalises the model through phased implementation, measurable indicators, coordinated oversight and independent evaluation. It concludes that Malaysian policy should measure protection not by reduced access alone, but by whether adolescents can participate online with competence, privacy, trusted support, effective remedies and meaningful institutional accountability.

Keywords: parental mediation, digital literacy, adolescent online safety, platform accountability, Malaysia

INTRODUCTION

Digital connectivity is no longer a peripheral feature of adolescent life in Malaysia. It is part of the infrastructure through which young people learn, communicate, build friendships, seek entertainment, participate in culture and form an understanding of themselves. Malaysia's wider digitalisation agenda has accelerated this transformation by treating connectivity, data and digital services as drivers of economic and social development (Economic Planning Unit, 2021). The result is a generation for whom the boundary between "online" and "offline" experience is increasingly artificial.

The scale of this transition is evident in national and child-focused data. The Department of Statistics Malaysia reported that individual internet use reached 98.3 per cent in 2025, while mobile-phone use reached 99.6 per cent (Department of Statistics Malaysia, 2026). Among internet-using children aged 12 to 17 surveyed in the Disrupting Harm study, 91 per cent used social media weekly, 90 per cent used instant-messaging services, 88 per cent watched video clips and 86 per cent used the internet for schoolwork (ECPAT International et al., 2022). These figures demonstrate that digital engagement is simultaneously social, recreational and educational. It cannot be treated as an optional leisure activity that can be removed without wider consequences for participation and development.

High connectivity does not, however, translate automatically into safe or critical use. Adolescents may be technically fluent in opening accounts, managing multiple profiles, navigating platform interfaces and producing content while remaining poorly equipped to assess commercial persuasion, privacy settings, manipulated information, harmful contact or the consequences of permanent data trails. Operational confidence can therefore conceal deficits in judgment, verification and help-seeking. The gap between knowing how to use a platform and knowing how to evaluate it is central to the concept of digital literacy.

Within the family, responsibility for closing that gap has often been placed on parents and caregivers. Parental mediation traditionally includes active discussion, restrictive rules and co-use. Contemporary practice also includes monitoring, technical controls, account supervision and modelling through the parent's own media behaviour (Clark, 2011; Livingstone & Helsper, 2008; Valkenburg et al., 1999). These strategies are not equivalent. A conversation that helps an adolescent evaluate a suspicious account builds a different capacity from a rule that blocks access. A device filter may prevent one exposure but does not necessarily teach the child how to respond when the same risk appears through another device, platform or peer network.

The Malaysian context reveals why a simple opposition between "strict" and "permissive" parenting is inadequate. Families differ in income, education, employment demands, language, household composition, access to devices and familiarity with changing platforms. The Disrupting Harm findings show a pronounced age-related gap among caregivers: only 44 per cent of caregivers aged 50 and above reported knowing how to report harmful social-media content, compared with 87 per cent of caregivers aged 29 and below. Rural children were also less likely than urban children to know how to assess the trustworthiness of a website or report harmful content (ECPAT International et al., 2022). The capacity to mediate digital life is therefore socially distributed rather than uniformly available.

Malaysia's legal environment has also changed. The Online Safety Act 2025 came into force on 1 January 2026 and imposes direct duties on licensed applications and content applications service providers. These duties include measures to mitigate harmful content, accessible safety tools, reporting and assistance mechanisms, protections for child users and an Online Safety Plan (Online Safety Act 2025, ss. 13–20). Section 18 is especially significant because it requires measures for the safe design and operation of services likely to be accessed by children, including controls on adult-child communication, features that extend use, access to children's personal information and personalised recommendation systems. The implementation architecture was further developed in 2026 through the Child Protection Code and Risk Mitigation Code, which took effect on 1 June 2026, and the Online Safety Plan Regulations, which took effect on 1 July 2026 (Parliament of Malaysia, 2026). Malaysia also began enforcing rules that prevent users below 16 from maintaining ordinary social-media accounts on regulated platforms, supported by age-verification measures (Bernama, 2026).

The regulatory turn is important because it recognises that families cannot redesign recommendation systems, moderate content at scale, prevent predatory account creation or alter the commercial incentives that reward prolonged engagement. Yet the introduction of age restrictions also risks renewing an access-centred approach in which safety is measured by whether adolescents are kept away from a service rather than whether the digital environment is safe, rights-respecting and developmentally appropriate. Restrictions may reduce some encounters, but they may also encourage circumvention, migration to less regulated spaces and reduced disclosure to adults.

This article argues that Malaysian adolescent online safety should move beyond parental control toward shared digital responsibility. Parents remain important, but their role should be enabling rather than substitutive: they should help adolescents interpret risk, preserve trust and seek assistance, while platforms, schools and regulators carry obligations that match their institutional capacities. Digital literacy must similarly be understood as a capability for safe participation, not merely a personal defence mechanism that transfers responsibility for systemic harm to the child.

The article makes three contributions. First, it integrates parental mediation theory with digital-literacy research, Malaysian administrative evidence and recent published consultations with children, showing that the relationship between skills and risk is not linear. Second, it evaluates the new platform duties and implementation instruments under the Online Safety Act 2025 against the limits of family-based protection and the emerging under-16 restriction. Third, it develops a Shared Digital Responsibility Model as a testable governance framework with defined responsibilities, implementation stages, monitoring indicators and review mechanisms.

Research objectives

The study has three objectives. The first is to explain how active, restrictive, supervisory and co-use forms of parental mediation affect adolescents' digital literacy, disclosure and exposure to online risk. The second is to assess whether Malaysia's current legal and policy framework distributes responsibility appropriately among families, schools, platforms and the state. The third is to formulate a practical model of shared digital responsibility that protects adolescents without treating exclusion from digital spaces as the principal measure of safety.

LITERATURE REVIEW

Parental mediation scholarship developed initially in relation to television and later expanded to internet and mobile-media use. The traditional framework distinguishes active mediation, restrictive mediation and co-use. Active mediation involves discussion, explanation and evaluation of media content. Restrictive mediation sets limits on time, content, contacts or activities. Co-use refers to parent and child using media together, although co-presence alone may involve little guidance. Later accounts added monitoring, supervision, technical mediation and parental modelling because interactive platforms are persistent, personalised and portable in ways that broadcast media were not (Clark, 2011; Livingstone & Helsper, 2008; Nikken & Schols, 2015; Valkenburg et al., 1999).

Clark's (2011) parental mediation theory for the digital age is particularly useful because it treats mediation as part of family communication and power rather than a set of isolated controls. Adolescents are not passive recipients of rules. They interpret rules, negotiate privacy, conceal activities, seek peer assistance and develop workarounds. The effect of a parental strategy therefore depends on the quality of the parent-child relationship and whether the adolescent considers the intervention legitimate. A technically effective restriction may be relationally counterproductive if it communicates distrust or creates fear that disclosure will lead to punishment.

Active mediation is generally associated with opportunities for explanation and reflection (Duerager & Livingstone, 2012; Livingstone et al., 2017). Parents may discuss why a request for personal information is unsafe, how manipulated images circulate, why an online claim should be verified or when contact should be blocked and reported. This form of mediation does not require the parent to know every application feature. It requires sufficient knowledge to ask meaningful questions, recognise uncertainty and respond without

immediately blaming the child. Active mediation can support critical literacy because it connects abstract safety principles to situations encountered by the adolescent.

Restrictive mediation can be appropriate where the risk is foreseeable and the adolescent's developmental capacity is limited. Rules on device use during sleep hours, location sharing, contact with unknown adults or financial transactions may reduce immediate harm. The difficulty arises when restriction becomes the dominant response. Rodríguez-de-Dios et al. (2018) found that restrictive mediation reduced adolescents' digital skills and thereby reduced both online risks and online opportunities. Their findings expose a central policy dilemma: lower exposure may appear successful while simultaneously weakening the skills needed for later independent use.

Digital literacy also has a dual relationship with risk. More digitally skilled adolescents may recognise privacy settings, verify information and use reporting tools more effectively. At the same time, they tend to participate in more online activities and therefore encounter more opportunities and more risks. Livingstone et al. (2017) describe this as an opportunity-risk relationship rather than a simple safety effect. Skills do not make exposure disappear; they affect how users interpret, manage and recover from it. A policy that treats any exposure as evidence of failure may therefore discourage the very participation through which competence develops.

Recent research also demonstrates that parental mediation is conditioned by family resources. Lou et al. (2024) found that parental education, economic circumstances and family structure were associated with different mediation practices and adolescents' digital literacy. Active mediation and supervision showed positive relationships with digital literacy, while the combined effects of mediation strategies could reduce some disadvantages associated with family background. The findings are not directly transferable to Malaysia, but they support an ecological view: family practice is affected by social context and cannot be evaluated as though all parents possess equal time, confidence and knowledge.

Malaysian research similarly indicates that parental digital competence is uneven. Hassan et al. (2023) found that age, internet experience and professional background among parents of adolescents in the Klang Valley were associated with differences in digital literacy, with skills remaining a particular weakness. The implication is important. Policy campaigns that merely instruct parents to monitor children may impose duties without providing the capability to perform them. A parent who does not understand disappearing messages, platform reporting pathways, recommender systems or account privacy cannot be expected to supply protection equivalent to that of a technically sophisticated parent.

The Ecological Systems Theory provides a broader framework. Bronfenbrenner (1979) explains development through nested systems, including the immediate family and school, relationships between those settings, wider institutional structures and the larger cultural environment. Applied to digital life, the adolescent's experience is shaped not only by household rules but also by school practices, peer norms, platform design, commercial incentives, law enforcement, data protection and national policy. A harmful encounter may begin with a platform recommendation, move through a peer network, produce conflict at home and eventually require school, counselling or regulatory intervention. No single layer can manage the entire chain.

A child-rights approach adds a further dimension. General Comment No. 25 of the United Nations Committee on the Rights of the Child requires digital policy to consider protection together with privacy, participation, access to information and the evolving capacities of the child (Committee on the Rights of the Child, 2021). The best interests of the child cannot be reduced to risk elimination. It requires transparent balancing among rights and attention to the views of children themselves. This approach is especially relevant to blanket age restrictions because a measure may reduce access to one category of service while also affecting learning, association, identity formation and access to support.

The literature consequently identifies four recurring problems. First, parental mediation is often treated as an individual family choice despite major differences in parental capability. Second, digital literacy is treated either as a universal solution or as a risk factor, when it is more accurately an enabling capability whose value depends on context. Third, restriction is easier to observe than resilience, leading policymakers to privilege measurable barriers over less visible capacities such as judgment, trust and recovery. Fourth, platform architecture is

frequently backgrounded even though recommendation, contact, visibility and engagement systems structure the risk environment before a parent becomes aware of it.

These problems support a shift from a parent-centred model to a responsibility-allocation model. The relevant question is not simply which parental strategy works best. It is which actor controls the source of a particular risk, which actor can prevent or mitigate it, and which actor should be accountable when protection fails.

METHODOLOGY

This study uses a doctrinal and evidence-informed socio-legal design. The doctrinal component analyses the Online Safety Act 2025, its emerging implementation instruments, the Personal Data Protection Act 2010 and the child-safety implications of Malaysia's regulated-platform framework. Particular attention is given to sections 13 to 20 of the Online Safety Act 2025, which establish provider duties concerning harmful-content risk mitigation, user guidelines, safety tools, reporting, assistance, child-user protection and Online Safety Plans. Official parliamentary records are used to verify the implementation status of the Child Protection Code, Risk Mitigation Code and Online Safety Plan Regulations as at July 2026.

Source identification was purposive and transparent rather than systematic. Contemporary materials published from 2011 to July 2026 were considered, while earlier works were retained where they supplied foundational theory or validated measurement concepts. The source pool comprised: (i) Malaysian legislation, subsidiary instruments and official parliamentary records; (ii) national statistics and institutional reports; (iii) Malaysia-specific or Malaysia-inclusive surveys and child-consultation reports; and (iv) peer-reviewed empirical studies examining parental mediation, digital literacy, online opportunities, risk, disclosure or family-resource inequalities.

Materials were included where they addressed adolescents or caregivers, explained a mechanism relevant to online safety, contained identifiable methods or official provenance, and contributed to at least one of four analytical outcomes: immediate exposure, digital competence, trust and disclosure, or access to institutional remedy. Studies concerning younger children were used only where the mediation mechanism was transferable and the age difference was expressly considered. Materials focused exclusively on adult users, purely technical cybersecurity without a child-safety dimension, or unsupported commentary were excluded. Non-Malaysian evidence was used to clarify mechanisms and possible implementation choices, not to establish Malaysian prevalence or causation.

Each source was charted against its jurisdiction, population, method, mediation strategy, risk mechanism, principal outcome and responsible institutional actor. The synthesis proceeded in four stages. First, it mapped access, common online activities and inequalities in safety capability. Second, it compared active, restrictive, supervisory and co-use strategies across the four analytical outcomes. Third, it examined whether the Online Safety Act 2025 addresses risks located beyond reasonable parental control. Fourth, it translated the combined doctrinal and empirical findings into the Shared Digital Responsibility Model and specified operational duties, indicators, evidence sources and review intervals.

Methodological triangulation was achieved by testing legal propositions against official statistics, caregiver and adolescent survey findings, and published child-consultation evidence. Consistent findings across different source types were treated as stronger support for the model, while differences in population, jurisdiction and research design were retained rather than collapsed into a single effect estimate. The study does not conduct a meta-analysis and does not infer causation from cross-sectional or qualitative evidence.

Scope and validation status

The study does not generate original interviews, surveys or platform data. Its empirical contribution lies in the structured synthesis of published Malaysian and comparative evidence, including direct consultations with children, rather than in a new primary dataset. The Shared Digital Responsibility Model is therefore presented as an evidence-informed and testable governance framework. Its practical value lies in converting broad calls for cooperation into observable responsibilities and measurable outcomes. Prospective validation should combine

adolescent and caregiver surveys, school-based pilots, platform audit data, complaint-outcome analysis and qualitative consultation, with results disaggregated where lawful and ethically appropriate by age, location, disability and household resources.

RESULTS AND DISCUSSION

The analysis produces six principal findings. First, Malaysian adolescents operate in a high-access environment in which online participation is embedded in education and relationships, making exclusion an incomplete safety strategy. Second, parental mediation affects not only exposure but also competence, trust and disclosure. Third, the benefits of mediation are distributed unequally because parental digital skills and family resources vary. Fourth, recent child-consultation evidence confirms that young users value online participation while also reporting that they may feel unsafe, unheard and left to manage harm without adequate institutional support. Fifth, the Online Safety Act 2025 creates a stronger basis for platform responsibility, particularly through child-safe design duties and implementation codes. Sixth, the under-16 restriction will be sustainable only if it is integrated with digital literacy, privacy safeguards, school support, accessible remedies and platform accountability.

High access, unequal safety capability

Malaysia's digital environment combines near-universal access with uneven safety capability. National connectivity statistics indicate that the principal divide is no longer simply between users and non-users. The more important divide concerns quality of access, privacy, skills, trusted support and the ability to obtain remedies.

The Disrupting Harm study illustrates this distinction. Approximately nine out of ten surveyed internet-using children engaged weekly in social media and instant messaging. Yet children in rural areas were less likely than urban children to know how to check whether a website could be trusted, with a reported difference of 43 per cent compared with 63 per cent. They were also less likely to know how to report harmful social-media content, at 58 per cent compared with 70 per cent (ECPAT International et al., 2022). These are not merely technical differences. They affect whether a child can recognise manipulation, interrupt an escalating interaction and reach a formal support channel.

The same pattern appears among caregivers. Older caregivers used the internet less frequently and reported weaker safety skills (ECPAT International et al., 2022; Hassan et al., 2023). A regulatory model that assumes every caregiver can configure controls, interpret platform settings and preserve digital evidence therefore risks deepening inequality. Better-resourced households can purchase devices with updated safety features, seek professional advice and devote time to co-use. Other households may rely on a shared phone, prepaid data or unsupervised access necessitated by work and schooling.

Multiple-account practices further complicate supervision. Adolescents may separate family-facing, peer-facing and interest-based identities. This can be part of ordinary social development and privacy management rather than evidence of wrongdoing. At the same time, fragmented identities make simple account inspection a poor proxy for safety. A parent may have access to one visible account while the adolescent's meaningful interactions occur elsewhere. The lesson is not that parents should intensify surveillance, but that durable protection depends on whether the adolescent can identify risk and seek help across accounts and platforms.

Digital inequality must therefore be understood as an inequality of protective capacity. Connectivity statistics can show that a household is online while concealing whether the child has a private device, a secure connection, age-appropriate settings, a trusted adult and the knowledge to report a problem. Policy should measure these dimensions separately.

Published stakeholder evidence and empirical grounding

Recent Malaysian consultation evidence supplies a direct stakeholder perspective that strengthens the socio-legal analysis. In 2025, 29 children from Sabah and Johor, including Deaf children, stateless children and children from urban-poor communities, were consulted on what would make the digital environment work better for them.

They described the internet as important for homework, family connection and exploration, but also reported feeling unsafe, not being heard and being left to manage online harms on their own (UNICEF Malaysia, 2026).

A wider UNICEF consultation involving more than 200 children across seven countries, including Malaysia, reached a closely related conclusion. Children identified stronger platform accountability, child-centred design, investment in digital literacy, equitable access and meaningful participation in policy design as priorities for the digital environment (Baghdasaryan & Vosloo, 2026). These findings do not substitute for original fieldwork by the present study, but they provide published empirical corroboration for the model's central claim that online safety requires coordinated responsibility rather than parental control alone.

The consultation findings also generate propositions capable of practical evaluation. A functioning shared-responsibility model should increase adolescents' knowledge of reporting pathways, access to a trusted adult, confidence that disclosure will not automatically result in total loss of access, and belief that platforms and institutions will respond. It should also reduce repeated exposure after a report, shorten referral and response times, and preserve beneficial participation for children whose education, disability support or social connection depends on digital access.

Why restriction can reduce disclosure

Restriction is attractive because it creates a clear rule and a visible compliance outcome. It may also be necessary for specific high-risk activities. However, evidence from Malaysia shows that punitive or access-threatening responses can reduce disclosure. Disrupting Harm reported that 36 per cent of caregivers said they would restrict their child's internet access if the child was bothered by something online (ECPAT International et al., 2022). The report warned that such a response may be perceived as punishment and could make children less likely to raise concerns. Earlier Malaysian focus-group findings similarly indicated that some adolescents concealed cyberbullying because they feared closer control or the loss of access (ECPAT International et al., 2022).

This dynamic can be explained through relational legitimacy. Adolescents are more likely to seek assistance when they expect to be heard, believed and supported. If disclosure triggers confiscation, moral judgment or public embarrassment, silence may appear safer. The parent then loses information precisely when intervention is most needed. A rule intended to protect the child can therefore create an incentive to hide the incident.

Restriction may also delay rather than resolve risk. A child barred from one platform may access it through a friend's device, a browser, a secondary account or a less regulated service. In such circumstances the family has less visibility, while the child may believe that seeking help would reveal the rule breach. This does not establish that age limits are always ineffective. It shows that enforcement must be accompanied by a non-punitive disclosure protocol.

A constructive protocol separates the harmful act of another user from the adolescent's possible breach of a family rule. The immediate priorities should be safety, evidence preservation, blocking, reporting and emotional support. Consequences for breaking a household rule, if any, should be addressed later and proportionately. This sequencing tells the adolescent that reporting harm will not automatically produce a greater loss.

Active mediation supports this approach because it treats online incidents as problems to interpret rather than proof of parental failure. Parents can rehearse responses before an incident occurs: stop communication, avoid forwarding content, capture necessary evidence, use platform tools and speak to a trusted adult. The goal is not to make the adolescent wholly self-reliant. It is to create a shared script that remains usable under stress.

Digital literacy as resilience rather than immunity

Digital literacy should not be framed as immunity from harm. No level of skill can guarantee safety in systems that use opaque ranking, behavioural profiling, viral distribution and automated persuasion. Skilled users may still encounter deceptive accounts, unwanted contact, scams or humiliating content. Moreover, adolescents' developmental stage can produce impulsivity, heightened sensitivity to peer approval and experimentation even where they understand an abstract rule.

The value of digital literacy lies in recognition, response and recovery. Recognition includes identifying suspicious requests, manipulated information, coercive communication, commercial influence and privacy consequences. Response includes changing settings, refusing contact, blocking, reporting, seeking help and preserving relevant evidence. Recovery includes restoring accounts, limiting further circulation, accessing support and learning without treating the incident as a permanent personal failure.

This capability-based understanding resolves the apparent contradiction that skilled adolescents experience both more opportunities and more risk (Livingstone et al., 2017). Greater participation creates more encounters. The policy objective should not be zero exposure, which is unrealistic, but reduced progression from exposure to harm. A digitally literate adolescent may encounter a suspicious message yet terminate the interaction before personal information is disclosed. Conventional prevalence measures may still count the encounter as a risk, while missing the successful protective response.

Parental mediation can strengthen this resilience when it is active, proportionate and developmentally responsive. Younger adolescents may require more direct supervision and narrower boundaries. Older adolescents should receive greater autonomy paired with clear expectations, periodic discussion and targeted intervention. The principle of evolving capacities supports this graduated approach (Committee on the Rights of the Child, 2021). Rules should change as competence, context and demonstrated responsibility change.

Schools are indispensable because they can reduce dependence on household resources. A national digital-literacy curriculum should move beyond password advice and generic warnings. It should include algorithmic awareness, commercial persuasion, source verification, privacy settings, respectful communication, reporting, evidence preservation and the social consequences of forwarding content. Scenario-based teaching is especially important because adolescents need practice applying principles to ambiguous situations.

The curriculum should also distinguish harmful content from disagreement and discomfort. Overbroad safety messages may encourage students to treat all unfamiliar views as dangerous, while narrow messages may overlook cumulative harassment or manipulative design. Critical digital literacy requires both caution and intellectual openness.

The Online Safety Act 2025 as a shift in responsibility

The Online Safety Act 2025 changes the legal allocation of responsibility by imposing direct duties on licensed providers. Section 13 requires measures to mitigate users' exposure to harmful content while providing that those measures should not unreasonably or disproportionately limit expression. Sections 14 to 17 require clear user guidelines, sufficient safety tools, reporting mechanisms and accessible assistance. These duties are important because they recognise that a reporting right is ineffective if the interface is difficult to locate, inaccessible or unresponsive.

Section 18 is the clearest child-specific provision. It requires providers to implement measures specified in the relevant code to ensure safe use by child users. The Act identifies design and operational measures that may prevent child access to suspected harmful content, limit adult communication with children, restrict features that increase or extend children's use, prevent adults from viewing children's personal information and control personalised recommendation systems for child users. Section 20 further requires an accessible and updated Online Safety Plan submitted to the Malaysian Communications and Multimedia Commission.

The institutional basis for implementation became more concrete in 2026. The Online Safety Committee established under section 5 provides a strategic forum for monitoring, research and recommendations, while MCMC remains the regulator responsible for compliance and enforcement. Parliament reported that the Child Protection Code and Risk Mitigation Code took effect on 1 June 2026 and that the Online Safety Plan Regulations took effect on 1 July 2026 (Parliament of Malaysia, 2026). The Shared Digital Responsibility Model can therefore be attached to institutions and instruments already in operation rather than requiring an entirely new regulatory body.

These provisions address risks located upstream of the family. A parent cannot independently know why a platform recommends a particular account, whether an adult has contacted thousands of child users, how quickly

harmful material is being replicated or whether an engagement feature is disproportionately affecting minors. Platforms possess the data and technical control needed to identify those patterns. The Act is therefore strongest where it aligns responsibility with control.

The practical challenge is auditability. A provider may state that it offers reporting tools or child settings, but the legal objective requires evidence of usability and effectiveness. Relevant indicators include the number of steps required to report, response time, recurrence after blocking, the default visibility of child accounts, the rate at which adult-child contact requests are interrupted and the treatment of recommendation pathways that repeatedly expose children to harmful themes. Online Safety Plans should explain not only what tools exist but also how providers test their operation.

Transparency must be balanced with privacy and security. Public disclosure should not reveal personal information or enable evasion of safety systems. Nevertheless, aggregate data, independent audits and regulator access are necessary if statutory duties are to produce more than formal compliance.

The Act should also be read together with data-protection principles. Age assurance and child-safety systems may process identity documents, facial information or behavioural signals. Data minimisation, purpose limitation, security and retention controls are therefore essential. The Personal Data Protection Act 2010 does not apply to the Federal and State Governments, which strengthens the case for explicit safeguards where public authorities participate in age-verification architecture or receive verification data (Department of Personal Data Protection, n.d.). Child protection should not create an unnecessarily permanent identity record of ordinary online participation.

The under-16 restriction: Protection, displacement and privacy

Malaysia's implementation of an under-16 social-media account restriction represents a strong precautionary response to public concern (Bernama, 2026). It may reduce exposure to specific features of major platforms and can place pressure on providers to improve age assurance. It also communicates that child safety is not solely a matter of parental vigilance.

The measure nevertheless has structural limits. First, age restrictions operate at the point of account access, while many risks occur through messaging, gaming, video, search, shared devices and smaller services. A narrow definition of social media may displace rather than remove risk. Second, adolescents who regard a restriction as illegitimate may falsify age information or migrate to spaces with weaker moderation. Third, account removal can affect educational channels, disability communities, peer support and creative participation. Fourth, identity-based verification can create privacy and exclusion concerns for users without reliable documentation.

UNICEF Malaysia (2025) has accordingly argued that age restrictions alone will not keep children safe. Its recommended approach combines platform redesign, child-friendly reporting, proportionate and privacy-preserving age assurance, stronger regulation, independent audits and meaningful transparency. This position is consistent with the Online Safety Act's own emphasis on design, tools and platform duties.

The restriction should therefore be treated as one layer in a wider system rather than the organising principle of child online safety. Its implementation should be evaluated against at least five questions. Has exposure to serious harm decreased, including after migration to other services? Are adolescents more or less likely to report problems? What data are collected for age assurance, by whom and for how long? Are vulnerable and undocumented children disproportionately excluded? Have platforms altered risky design features for all child users, or merely removed identified accounts?

A policy that answers only how many underage accounts were closed may overstate success. Effective protection requires outcome measures that include displacement, privacy, disclosure and access to beneficial services.

A shared digital responsibility model

The findings support a Shared Digital Responsibility Model based on five actors: adolescents, caregivers, schools, platforms and regulators. The model rejects both extremes. It does not place full responsibility on children to

protect themselves, and it does not assume that parents or the state can eliminate risk through surveillance and exclusion. Responsibility is assigned according to control, capability and proximity to harm.

Adolescents have participatory responsibilities appropriate to their evolving capacities (Committee on the Rights of the Child, 2021). These include using safety tools, respecting others, protecting account credentials, pausing suspicious interactions and seeking assistance. These are educational expectations, not grounds for blaming a child when another person or a platform causes harm. An adolescent's mistake should not erase the duties of adults, institutions or service providers.

Caregivers have a duty of enabling mediation. They should establish developmentally appropriate boundaries, maintain open communication, model responsible use and respond to disclosure without immediate punishment. Their role includes helping the adolescent interpret experiences and reach formal assistance. Caregiver duties must be supported through accessible training rather than assumed.

Schools have an equalising responsibility. They should provide structured digital-literacy education, identify trusted reporting persons, maintain protocols for incidents that cross home and school boundaries and support parents with practical guidance. Schools should not require invasive access to students' private accounts as a routine disciplinary practice. Their intervention should be necessary, proportionate and connected to student welfare.

Platforms carry the strongest preventive responsibility where harm is shaped by service design, recommendation, visibility, contact or monetisation. They should implement child-safe defaults, limit unnecessary data collection, reduce unsolicited adult contact, test reporting tools with children, interrupt repeated harmful recommendations and provide rapid escalation for serious cases. Safety should be assessed through outcomes rather than the mere presence of features.

Regulators and the state have responsibilities of standard-setting, auditing, coordination and remedy. They should define measurable expectations under the Online Safety Act 2025, require meaningful Online Safety Plans, publish aggregate enforcement data and establish child-friendly complaint pathways. They should also ensure that age-assurance measures are proportionate and privacy-preserving, and that digital-literacy support reaches families with lower resources.

The model is operationalised through four principles. The first is prevention at source: the actor controlling the relevant system should address foreseeable risk before transferring responsibility downstream. The second is enabling autonomy: interventions should increase adolescents' ability to participate safely rather than preserve permanent dependence. The third is trusted disclosure: no safety system is effective if adolescents expect blame, humiliation or automatic exclusion when they seek help. The fourth is auditable accountability: every institutional duty should be linked to evidence capable of independent review.

Operational framework, coordination and evaluation

Table 1: Shared Digital Responsibility Model

Actor	Core responsibility	Operational measure	Illustrative accountability indicator
Adolescents	Use skills and seek assistance in line with evolving capacities	Scenario-based safety rehearsals, reporting, account security and respectful participation	Correct completion of safety scenarios; knowledge of reporting tools; access to a trusted adult
Caregivers	Provide enabling mediation and proportionate boundaries	Regular discussion, non-punitive disclosure protocol, targeted supervision and referral knowledge	Caregiver module completion; adolescent willingness to disclose; confidence in incident response

Actor	Core responsibility	Operational measure	Illustrative accountability indicator
Schools	Equalise access to digital literacy and support	Age-progressive curriculum, trained focal person, practical assessment and incident-referral protocol	Curriculum coverage; scenario-assessment results; time from disclosure to appropriate referral
Platforms	Prevent foreseeable harm through safe design and operation	Child-safe defaults, adult-contact limits, recommender testing, accessible reporting, appeals and error correction	Median response time; repeat-exposure rate; reporting usability; appeal outcomes; age-assurance error rate
Regulators and state	Set standards, coordinate actors, audit compliance, protect privacy and provide remedies	Guidance, standardised Online Safety Plans, independent audits, aggregate transparency and privacy review	Audit coverage; compliance findings; remedy outcomes; annual review completion; inequality impact

The Shared Digital Responsibility Model can be implemented through the existing architecture of the Online Safety Act 2025. The Online Safety Committee should coordinate cross-sector priorities and advise on a national indicator framework; MCMC should translate statutory duties into auditable provider requirements; the Ministry of Education should oversee curriculum and school incident protocols; the ministry responsible for child and family welfare should coordinate protection and referral services; the data-protection authority should scrutinise age-assurance data practices; and platforms should supply standardised compliance data. A standing adolescent advisory panel should participate in the design and review of reporting tools, guidance and outcome measures.

Implementation should proceed in three stages. During the first six months, responsible bodies should establish a baseline, agree common definitions, map existing reporting and referral routes, and co-design indicators with adolescents, caregivers, educators and disability advocates. Between months six and eighteen, the framework should be piloted through selected schools, caregiver-training sites and regulated platforms. From months eighteen to twenty-four, an independent evaluation should compare baseline and pilot outcomes, identify unequal effects and recommend national scaling. Thereafter, the indicator framework and provider obligations should be reviewed annually.

For platforms, Online Safety Plans should use a standard reporting template covering child-account defaults, adult-child contact controls, recommender-system testing, time-extending features, reporting accessibility, median response time, repeat exposure after blocking or reporting, appeal outcomes and age-assurance error correction. Independent audits should test actual user pathways, including accessibility for children with disabilities, rather than reviewing policy documents alone. Public reporting may use aggregate data, while MCMC should retain access to sufficiently detailed evidence for enforcement.

For caregivers and schools, implementation should focus on capability rather than surveillance. Caregiver modules should teach non-punitive disclosure, evidence preservation, blocking, reporting and referral. Schools should deliver age-progressive digital-literacy instruction, appoint trained focal persons and adopt protocols for incidents that cross home, school and platform boundaries. Scenario-based assessment should test whether students can recognise manipulation, use reporting tools and identify a trusted adult, rather than merely recall safety slogans.

Coordination should follow a documented referral protocol. Routine platform disputes should move through accessible platform review and appeal processes. Reports indicating immediate danger, exploitative contact or serious child-protection concerns should be escalated promptly to the appropriate specialist authority. The referring institution should record the action taken, transfer only necessary information and provide understandable feedback to the child and caregiver. Clear ownership at each stage is necessary to prevent reports from being passed repeatedly between institutions.

Evaluation should combine administrative and experiential measures. Administrative measures include reporting and response times, repeat-exposure rates, appeal outcomes, school referral times, audit findings and compliance

directions. Experiential measures include adolescent knowledge, willingness to disclose, access to a trusted adult, perceived fairness, caregiver confidence and continued access to beneficial digital participation. Baseline, twelve-month and twenty-four-month assessments should be complemented by periodic qualitative consultation so that apparently positive compliance figures are tested against children's actual experiences.

Age-assurance governance requires a separate privacy evaluation. Providers and verification vendors should identify the data used, purpose, retention period, security controls, error rates and appeal route. The least intrusive reasonably effective method should be preferred, and verification data should not be repurposed for advertising, profiling or unrelated identity tracking. Evaluation should consider both safety outcomes and exclusion risks affecting undocumented children, children with disabilities and users who are wrongly classified.

Table 2: Phased Implementation and Evaluation Protocol

Stage and timing	Coordinating actors	Required actions	Review evidence
Baseline and co-design 0–6 months	Online Safety Committee, MCMC, education and child-welfare authorities, data-protection authority, platforms and adolescent representatives	Agree definitions and referral routes; map existing capacity; establish baseline; co-design indicators and accessible reporting pathways	Baseline dashboard; documented governance map; privacy and accessibility review; consultation record
Pilot and capacity building 6–18 months	MCMC, selected platforms, schools, caregiver-training partners and independent evaluators	Pilot platform metrics, school curriculum, focal-person protocols, caregiver modules and age-assurance safeguards	Six- and twelve-month monitoring reports; usability tests; audit logs; participant feedback
Independent evaluation and scaling 18–24 months	Independent evaluator reporting to the Online Safety Committee and MCMC	Compare baseline and pilot outcomes; assess unequal effects; verify compliance evidence; recommend revision and scaling	Twenty-four-month mixed-method evaluation; public summary; corrective action plan
Continuous review Annually	Online Safety Committee and MCMC with cross-sector and adolescent participation	Review indicators, emerging risks, complaint outcomes, privacy impacts and implementation gaps	Annual public report; updated guidance; provider improvement commitments; enforcement follow-up

CONCLUSION

This article examined the relationship between parental mediation, digital literacy and online risk among Malaysian adolescents at a moment of significant regulatory change. The combined doctrinal, survey and child-consultation evidence shows that adolescent digital life cannot be governed effectively through a single strategy. Restriction can reduce immediate exposure, but may also reduce skills, opportunities and disclosure. Digital literacy improves recognition and response, but does not immunise adolescents from systems designed to maximise attention, contact and data extraction. Active mediation and proportionate supervision are more compatible with resilience, although their effectiveness depends on parental competence, time and trust.

Malaysia's Online Safety Act 2025 is therefore an important development. By imposing duties concerning harmful-content mitigation, safety tools, reporting, assistance, child-safe design and personalised recommendation systems, it acknowledges that platforms control risks that families cannot. The Child Protection Code, Risk Mitigation Code, Online Safety Plan Regulations and Online Safety Committee provide an emerging institutional structure through which those duties can be specified, coordinated and reviewed.

The under-16 social-media restriction should be understood within that wider framework. It may reduce particular forms of exposure and accelerate age-assurance reform, but it cannot be the sole measure of success. Risk can migrate to other services, adolescents can circumvent restrictions and identity-based verification can create privacy and exclusion costs. Evaluation must therefore consider displacement, disclosure, beneficial participation, error correction and data governance.

The proposed Shared Digital Responsibility Model offers a more balanced allocation of duties and a practical evaluation framework. Adolescents should develop skills without being blamed for systemic harm. Caregivers should provide enabling mediation without being expected to replace technical regulation. Schools should equalise capabilities. Platforms should prevent foreseeable harm at source. Regulators should define and enforce measurable standards. The model's staged implementation, institutional allocation, audit sources and review intervals make these responsibilities capable of empirical testing rather than leaving them as a general aspiration.

The central policy shift is from controlling adolescent access to governing the conditions of adolescent participation. A safer digital environment is not one in which children are merely absent. It is one in which their presence is anticipated, their views inform design, their families receive meaningful support and institutions can demonstrate that preventable harm is reduced, reports lead to timely action and beneficial participation remains protected.

ETHICAL APPROVAL

Not applicable. This doctrinal and socio-legal study did not involve human participants, animals, interviews or the collection of personal data.

DATA AVAILABILITY

No new primary dataset was generated for this study. The analysis relies on publicly available legislation, official statistics, parliamentary records, institutional reports, published child-consultation materials and peer-reviewed scholarship identified in the reference list.

ACKNOWLEDGEMENTS

The authors wish to thank Universiti Kebangsaan Malaysia for supporting this study.

FUNDING

This study was funded by Universiti Kebangsaan Malaysia research grant GUP-2025-016.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. Baghdasaryan, B., & Vosloo, S. (2026). Children's best interests in digital policy and practice: From the voices of children, youth and experts around the world. UNICEF.
2. Bernama. (2026, June 1). Malaysia begins enforcing social media age verification rules. <https://bernama.com/en/news.php?id=2563426>
3. Bronfenbrenner, U. (1979). The ecology of human development: Experiments by nature and design. Harvard University Press.
4. Clark, L. S. (2011). Parental mediation theory for the digital age. *Communication Theory*, 21(4), 323–343. <https://doi.org/10.1111/j.1468-2885.2011.01391.x>
5. Committee on the Rights of the Child. (2021). General comment No. 25 (2021) on children's rights in relation to the digital environment (CRC/C/GC/25). United Nations.
6. Department of Personal Data Protection. (n.d.). Application and non-application of the Personal Data Protection Act 2010. <https://www.pdp.gov.my/ppdpv1/en/akta/application-and-non-application-of-the-act/>
7. Department of Statistics Malaysia. (2026). ICT use and access by individuals and households survey report, 2025. <https://www.dosm.gov.my/portal-main/release-content/ict-use-and-access-by-individuals-and-households-survey-report-2025>
8. Duerager, A., & Livingstone, S. (2012). How can parents support children's internet safety? EU Kids

Online.

9. Economic Planning Unit. (2021). Malaysia digital economy blueprint. Prime Minister's Department.
10. ECPAT International, INTERPOL, & UNICEF Office of Research – Innocenti. (2022). Disrupting harm in Malaysia: Evidence on online child sexual exploitation and abuse. UNICEF.
11. Hassan, H. M., Mohamad Salleh, M. A., & Ahmad, A. L. (2023). Digital literacy to digital fluency among parents in mediating adolescents' online safety. *International Online Journal of Language, Communication, and Humanities*, 6(2), 17–32. <https://doi.org/10.47254/6bb1tc67>
12. Livingstone, S., & Helsper, E. J. (2008). Parental mediation of children's internet use. *Journal of Broadcasting & Electronic Media*, 52(4), 581–599. <https://doi.org/10.1080/08838150802437396>
13. Livingstone, S., Mascheroni, G., & Staksrud, E. (2018). European research on children's internet use: Assessing the past and anticipating the future. *New Media & Society*, 20(3), 1103–1122. <https://doi.org/10.1177/1461444816685930>
14. Livingstone, S., Ólafsson, K., Helsper, E. J., Lupiáñez-Villanueva, F., Veltri, G. A., & Folkvord, F. (2017). Maximizing opportunities and minimizing risks for children online: The role of digital skills in emerging strategies of parental mediation. *Journal of Communication*, 67(1), 82–105. <https://doi.org/10.1111/jcom.12277>
15. Lou, J., Wang, M., Xie, X., Wang, F., Zhou, X., Lu, J., & Zhu, H. (2024). The association between family socio-demographic factors, parental mediation and adolescents' digital literacy: A cross-sectional study. *BMC Public Health*, 24, Article 2932. <https://doi.org/10.1186/s12889-024-20284-4>
16. Nikken, P., & Schols, M. (2015). How and why parents guide the media use of young children. *Journal of Child and Family Studies*, 24, 3423–3435. <https://doi.org/10.1007/s10826-015-0144-4>
17. Online Safety Act 2025 (Act 866) (Malaysia).
18. Parliament of Malaysia. (2026, July 2). Digital Hansard: Dewan Rakyat sitting, 2 July 2026. <https://hansard.parlimen.gov.my/hansard/dewan-rakyat/2026-07-02>
19. Personal Data Protection Act 2010 (Act 709) (Malaysia).
20. Rodríguez-de-Dios, I., van Oosten, J. M. F., & Igartua, J. J. (2018). A study of the relationship between parental mediation and adolescents' digital skills, online risks and online opportunities. *Computers in Human Behavior*, 82, 186–198. <https://doi.org/10.1016/j.chb.2018.01.012>
21. Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., Livingstone, S., & Hasebrink, U. (2020). EU Kids Online 2020: Survey results from 19 countries. *EU Kids Online*. <https://doi.org/10.21953/lse.47fdeqj01ofo>
22. Third, A., Bellerose, D., Dawkins, U., Keltie, E., & Pihl, K. (2014). Children's rights in the digital age: A download from children around the world. Young and Well Cooperative Research Centre and UNICEF.
23. UNICEF Malaysia. (2020). Our lives online: Use of social media by children and adolescents in East Asia—Opportunities, risks and harms. <https://www.unicef.org/malaysia/reports/our-lives-online>
24. UNICEF Malaysia. (2025, December 15). Age restrictions alone won't keep children safe online, says UNICEF. <https://www.unicef.org/malaysia/press-releases/age-restrictions-alone-wont-keep-children-safe-online-says-unicef>
25. UNICEF Malaysia. (2026). Through the lens of children in Sabah and Johor: A child-friendly report. <https://www.unicef.org/malaysia/documents/through-lens-children-sabah-and-johor-child-friendly-report>
26. UNICEF Office of Research – Innocenti. (2025). Childhood in a digital world. UNICEF.
27. Valkenburg, P. M., Krcmar, M., Peeters, A. L., & Marseille, N. M. (1999). Developing a scale to assess three styles of television mediation: Instructive mediation, restrictive mediation, and social covieing. *Journal of Broadcasting & Electronic Media*, 43(1), 52–66. <https://doi.org/10.1080/08838159909364474>