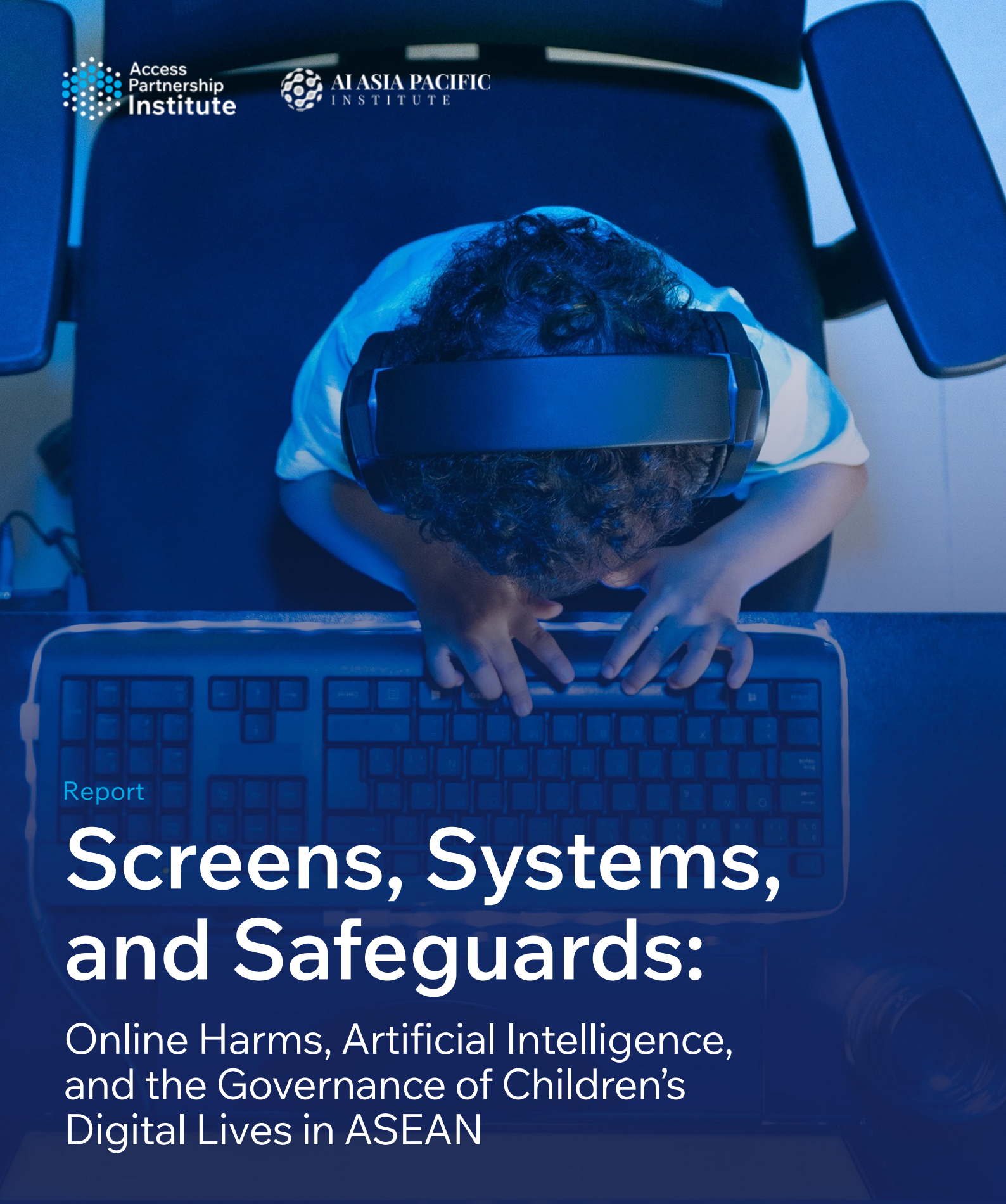

Report

Screens, Systems, and Safeguards:

Online Harms, Artificial Intelligence, and the Governance of Children's Digital Lives in ASEAN

MAY 2026

Contents

Acknowledgements

We would like to thank the following for their contributions to the paper:

Access Partnership

Jonathan Gonzalez
Associate Director, Public Sector Advisory

Lim May-Ann
Director, Multilateral Relations, Data Policy, and Partnerships

Riccardo Gualandi
Head of Design

AI Asia Pacific Institute

Kelly Forbes
President

Dr Peter Brimble,
Vice President

Dr Wendy Bonython
Secretary

Hamzah Bin Zaid
Associate

Meghna Ravi
Research Intern

Definitions

Agentic AI	Autonomous systems capable of setting their own goals, planning, and executing multi-step actions with minimal human supervision . Unlike traditional AI, Agentic AI does not require explicit direction at each step.
Artificial Intelligence (AI)	Digital systems that perform a wide range of activities, including generating content, interpreting information, solving problems, adapting to feedback, and interacting with environments. AI operates with varying degrees of autonomy and generality.
ASEAN (Association of Southeast Asian Nations)	A regional intergovernmental organisation comprising Brunei, Cambodia, Indonesia, Laos, Malaysia, Myanmar, the Philippines, Singapore, Thailand, and Vietnam.
Cyberbullying	The practice of bullying a person through online services or platforms, typically via messages or content that is intimidating, mocking, or threatening in nature.
Deepfake	A type of synthetic media generated using AI to produce convincingly fabricated images, videos, or audio recordings.
European Union (EU)	A supranational political and economic union of 27 member states, located primarily in Europe.
GenAI (Generative AI)	A subset of artificial intelligence capable of generating text, images, or other media in response to prompts, using pre-trained transformer models typically trained on large-scale datasets.
Institute of Higher Learning (IHL)	Post-secondary educational institutions – including universities, polytechnics, and technical or arts institutes – authorised to grant diplomas, degrees, or equivalent qualifications.
Minor	A person under the age of full legal responsibility, typically 18 years in most jurisdictions.
Non-Governmental Organisation (NGO)	An organisation operating independently of government at local, national, or international level to address specific issues or policy matters.
Statutory Board	An autonomous government agency established by a specific Act of Parliament to perform specialised public functions or manage public enterprises.
Social Media	Internet-based applications and websites that allow users to create, share, and exchange information, ideas, opinions, and multimedia content within virtual communities and networks.

AI, Social Media, and the New Risk Landscape for Minors

Artificial intelligence – and the rapid diffusion of generative AI, agentic AI, and algorithmic recommender systems – has reshaped the digital environment in ways that now sit squarely within the domain of online safety and digital harms governance.

The AI Asia Pacific Institute (AIAPI)-Netsafe discussion paper *AI and Online Safety: Emerging Risks and Opportunities* frames AI as a dual-use capability: a tool that enables new forms of creativity and productivity while simultaneously amplifying a broad spectrum of online harms, from misinformation and scams to cyberbullying, harassment, and child sexual exploitation and abuse (CSEA).[1]

The harms are not new. What has changed is the scale at which AI and social media can amplify harmful dynamics once content becomes easier to produce, distribute, and personalise.

This paper builds on that framing. It contends that AI's most consequential impact on minors is less about isolated "bad content" and more about system-level shifts: how platforms curate engagement, how synthetic media destabilises trust and consent, and how safety interventions are becoming central components of AI governance.

The regulatory salience of these shifts rests on a structural reality: online spaces are no longer neutral conduits for information. They are engineered environments in which AI systems increasingly decide what is seen, when it is seen, and in what social context it is encountered. The AIAPI–Netsafe paper identifies a set of AI "capability upgrades" – personalisation, and improvements in content quality, accessibility, and speed of generation – that matter because they alter the economics and mechanics of digital interaction.

As a result, minors' online experiences are progressively shaped by algorithmic inferences about preferences, vulnerabilities, and likely responses, not by chronological feeds or deliberate search entries alone.

1.1. What Social Media and AI Are Doing to Young People

These engagement-optimised systems can produce compulsive or difficult-to-control usage patterns, particularly among children and adolescents still developing self-regulation capacities and risk appraisal skills.

The American Psychological Association (APA) has cautioned that social media's effects are not uniformly positive or negative but depend on adolescents'

[1] AI Asia Pacific Institute & Netsafe. (2024). AI and online safety: Emerging risks and opportunities. https://aiasiapacific.org/our-work/ai-and-online-safety-emerging-risks-and-opportunities

characteristics, contexts, and the design features of the platforms themselves (American Psychological Association, 2023).[2]

The APA flags specific design and interaction features – recommended content, unrestricted time limits, and endless scrolling – as requiring age-appropriate, contextual tailoring rather than being inherited from adult product design.

Public health-oriented evidence has reinforced these concerns. A widely cited longitudinal study in *JAMA Psychiatry* (Riehm et al., 2019) found that adolescents spending more than three hours per day on social media "may be at heightened risk" for mental health problems, particularly internalising problems.

A more recent cohort study published in *JAMA Network Open* (Nagata et al., 2025) tracked approximately 12,000 children from ages 9-10 over three years. Over that period, average daily social media use rose from 7 minutes to 73 minutes and mean depressive symptom scores increased by 35 per cent – with within-person analysis suggesting that increased social media use preceded the rise in depressive symptoms, rather than the reverse.

Additionally, as AI becomes embedded in gaming to create more immersive experiences, personalisation drives deeper emotional attachment between minors and the characters or environments they interact with.

The U.S. Surgeon General's Advisory similarly argues that current evidence does not permit the conclusion that social media is sufficiently safe for children and adolescents, pointing to findings linking heavy daily use with elevated depression and anxiety symptoms (U.S. Department of Health and Human Services, 2023).[3]

These sources converge on a premise that matters for regulation: the risk profile of digital platforms is shaped by the engagement and recommendation machinery that can repeatedly steer minors towards harmful material or reinforce maladaptive comparison and attention patterns.

"AI and social media can pose real risks to minors' attention, identity, and mental health. Rules and platform controls, while useful, are often blunt instruments that can be bypassed, particularly by increasingly tech-savvy minors."

– **Amrin Amin**, Head, Corporate Development, Temasek Foundation (Former Senior Parliamentary Secretary, Ministry of Health)

These risks are recognised across both clinical and policy communities. Amrin Amin, formerly Senior Parliamentary Secretary at Singapore's Ministry of Health, observes that AI and social media can pose real risks to minors' attention, identity, and mental health. He adds that rules and platform controls, while useful, are often blunt instruments that can be bypassed, particularly by increasingly tech-savvy minors.[4]

[2] American Psychological Association. (2023). Health advisory on social media use in adolescence. www.apa.org/topics/social-media-internet/health-advisory-adolescent-social-media-use.pdf
[3] U.S. Department of Health and Human Services. (2023). Social media and youth mental health: The U.S. Surgeon General's advisory. www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf
[4] See section Annex: Contributors.

1.2. How Digital Harms Reach Minors

The digital harms landscape for minors is multi-dimensional – spanning interpersonal harms (cyberbullying, harassment, grooming), informational harms (misinformation and disinformation), consumer harms (exploitation, unwanted spending), and privacy harms (data misuse, unwanted contact, and exposure through weak defaults).

Category of Harm	Examples	AI's Role in Amplification
Child Sexual Exploitation and Abuse	CSAM; grooming; sextortion	AI-generated or manipulated CSAM; AI-facilitated grooming tactics; AI-generated imagery used in sextortion
Violent or Graphic Content	Violent imagery; graphic videos; incitement of violence	AI-generated violent or graphic content produced for extremist or disinformation purposes
Extremism	Violent extremism; terrorism	AI-generated or AI-translated propaganda; personalised propaganda messaging; AI-generated content that evades detection
Harm to Health and Well-being	Harmful social media trends; self-harm incitement; eating disorders and body dysmorphia	AI-facilitated harmful trends; hallucinated misinformation; chatbots providing contextually inappropriate guidance; human–AI emotional entanglement
Indecent and Obscene Content	Pornography; non-consensual intimate imagery (NCII); hate; abuse and coercion	Deepfake pornography and non-consensual intimate imagery (NCII); use of deepfakes in blackmail and extortion
Hate and Discrimination	Online hate speech; discriminatory impacts of AI classification systems	AI-generated content reinforcing stereotypes or misrepresenting sensitive historical events; AI-facilitated hate speech and discrimination
Cyberbullying and Harassment	Abusive communications; exclusionary behaviour; stalking	AI-generated abusive content; deepfake-enabled harmful pranks (e.g. swatting); NCII deepfakes used to harass or coerce; AI-facilitated stalking
Misinformation and Disinformation	Fake news; troll farms; spam bots; information echo chambers	Deepfakes; AI-facilitated troll farms; micro-targeted personalised content; AI chatbot "nudging"; hallucinations
Scams	Phishing; fraudulent schemes	AI-enabled voice cloning and deepfake impersonation; improved quality and personalisation of phishing; AI-facilitated scam websites and accounts

Some of these harms are persistent and measurable even before the added complication of generative AI. The World Health Organization (WHO) reports that cyberbullying has increased in recent years, with approximately 15 per cent of school-aged children in Europe (roughly one in six) experiencing cyberbullying – up from 13 per cent in 2018 (WHO Regional Office for Europe, 2024).[5]

The study warns that intensifying digitalisation can have profound impacts on young lives. For AI governance, the relevance is straightforward: algorithmic systems can magnify or accelerate these harms. AI can generate abusive messages at scale, automate impersonation, and produce psychologically targeted content aimed at particular individuals or peer groups.

Generative tools also lower the capability threshold for producing and disseminating synthetic or manipulated media – particularly deepfakes, which have become a cross-cutting risk multiplier. The AIAPI–Netsafe paper describes deepfakes as spanning two dominant harm domains: "falsity" (distorting reality) and "nudity" (creating sexualised content). Advances have made deepfakes more accessible, harder to detect, and harder to police – especially where models can be run locally on consumer hardware.

These harms are not speculative. The FBI has warned that it continues to receive reports – including involving minor victims – of benign photographs or videos being manipulated into explicit content and circulated for harassment or sextortion (FBI, 2023).[6]

Research by Sensity AI and Home Security Heroes has consistently found that women comprise approximately 99 per cent of targets of non-consensual deepfake pornography, which itself represents the overwhelming majority (96-98 per cent) of all deepfake content online (Ajder et al., 2019;[7] Home Security Heroes, 2023).[8]

At the extreme end, child-protection organisations have documented emerging risks from AI-generated child sexual abuse material. The Internet Watch Foundation (IWF) has warned that AI-generated child sexual abuse imagery poses a significant and growing threat, including the large-scale creation of new abusive imagery from existing material (IWF, 2023).[9]

The policy implication is clear: AI-generated content scales not only legitimate creativity but also production of harmful materials. Legal frameworks designed around the assumption that abuse content is rare or difficult to produce must adapt to an environment in which creation is cheap and automated.

[5] WHO Regional Office for Europe. (2024). A focus on adolescent peer violence and bullying in Europe, central Asia and Canada: Health Behaviour in School-aged Children international report from the 2021/2022 survey. www.who.int/europe/news/item/27-03-2024-one-in-six-school-aged-children-experiences-cyberbullying--finds-new-who-europe-study
[6] Federal Bureau of Investigation. (2023). Malicious actors manipulating photos and videos to create explicit content and sextortion schemes. Public Service Announcement. www.ic3.gov/PSA/2023/PSA230605
[7] Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). The state of deepfakes: Landscape, threats, and impact. Deeptrace (now Sensity AI). https://sensity.ai/reports/
[8] Home Security Heroes. (2023). 2023 state of deepfakes: Realities, threats, and impact. www.securityhero.io/state-of-deepfakes
[9] Internet Watch Foundation. (2023). How AI is being abused to create child sexual abuse imagery. www.iwf.org.uk/news-media/news/worst-nightmares-come-true-as-predators-are-able-to-make-thousands-of-new-ai-images-of-real-child-victims

1.3. Why Existing Rules Are Not Enough

Regulatory activity in the EU and China, among other jurisdictions, can be read as a response to the widening gap between the harms that emerge in AI-mediated environments and the protections that traditional platform self-governance can reliably deliver. At least three governance constraints stand out.

"When public sentiment reaches a tipping point, regulation tends to focus on what technology is doing to children rather than on children's developmental needs and perspectives."

– **Dr Nikhila Natarajan**, Adjunct Professor, School of Communication and Information, Rutgers, The State University of New Jersey

First, enforcement and accountability are strained by the international nature of platform services and the growing availability of decentralised tools. The AIAPI–Netsafe paper argues that the cross-border character of technological development and online services creates strong scope for international coordination – harms can scale across borders even when legal authority does not. It also warns that as AI tools become deployable on consumer-grade hardware, interventions relying on a small number of chokepoints (cloud providers, model developers) may lose effectiveness.

Second, online harms are often driven by misuse rather than intended use. The AIAPI–Netsafe paper notes that many generative AI risks originate from human behaviour, and that an overly narrow focus on organisational governance can underemphasise end users as both perpetrators and victims.

This matters especially for minors, who are both disproportionately exposed to harms and embedded in peer dynamics where low-friction tools can be turned to humiliation, coercion, or social control.

"Effective regulation should move towards child-centred, risk-based frameworks prioritising prevention, safety-by-design, accountability, and children's active involvement in decision-making."

– **Maryam Ehsani**, CEO, Child Safe Me

Third, incentives within attention-based business models can conflict with child safety imperatives. While industry-led tools such as parental controls and reporting mechanisms serve a purpose, policymakers are increasingly treating default settings, interface design, and recommender objectives as decisions too consequential to leave solely to commercial discretion.

These structural constraints are compounded by a more fundamental misalignment in how regulation is designed. As Dr Nikhila Natarajan observes, the pattern is well-established: when public sentiment reaches a tipping point, regulation tends to focus

on what technology is doing to children rather than on children's developmental needs and perspectives.[10]

For Dr Natarajan, past legislative efforts such as the V-chip and COPPA illustrate the limitations of approaches that target specific content or data collection practices without addressing the broader industry structures that produce harm.

The implication for current debates is that even well-intentioned interventions risk reproducing this pattern if they remain anchored in platform categories rather than in a developmental understanding of how minors actually engage with digital environments.

This direction is consistent with the practitioner consensus emerging from Maryam Ehsani's work in this area. She makes the case that effective regulation should move towards child-centred, risk-based frameworks prioritising prevention, safety-by-design, accountability, and children's active involvement in decision-making.[11]

The EU's *Digital Services Act (DSA) Guidelines* on the protection of minors, which provides detailed recommendations on recommender adaptations, limits on engagement-based signals, and restrictions on manipulative design, exemplify the shift from voluntary "good practice" to quasi-prescriptive safety design expectations (European Commission, 2025).[12]

[10] See section Annex: Contributors.
[11] See section Annex: Contributors.
[12] European Commission. (2025). Guidelines on the protection of minors under Article 28 of the Digital Services Act. www.agcom.it/sites/default/files/media/allegato/2025/EU%20Guidelines_for_minors_online_28%20DSA.pdf

Five Jurisdictions, Five Experiments: A Comparative Analysis

Section 1 framed the AI safety governance challenge as structural: minors' digital experiences are shaped not only by "harmful content" but by platform architectures and recommendation algorithms that scale exposure, dissemination, and risk.

In response, policymakers are experimenting with narrower and more interventionist regulatory responses – restricting minors' access to certain platform categories (typically social media), alongside broader "systems-based" safety obligations that regulate platform design and operational practices.

This section compares five jurisdictions that have recently accelerated their policy responses. It maps the measures each country has initiated; compares their legal form and enforcement logic; and identifies convergences and divergences, surfacing gaps, blind spots, and side-effects that could shape implementation outcomes.

2.1. Overview of Recent Measures

Jurisdiction	Primary Instrument	Status (as of March 2026)	Age Threshold	Core Scope	Primary Enforcement Lever
Australia	Social media minimum age obligation under the *Online Safety Amendment (Social Media Minimum Age) Act 2024*	Commenced 10 Dec 2025	Under 16	Designated "age-restricted" social media platforms	Platform duty to take "reasonable steps" to prevent under-16 accounts; civil penalties up to ~A$49.5m (eSafety Commissioner)
China	Minors' online protection framework (Minors Protection in Cyberspace + "Minors Mode" guidance + content classification measures)	In force / rolling measures (core regs effective Jan 2024; further measures issued late 2025, effective 2026)	No single ban	Online products/services used by minors; Minors Modes; classification of online information affecting minors	Mandatory platform/system controls; classification and labelling rules effective March 2026 (China Law Translate)
Spain	Proposed ban on social media access for under-16s	Announced; legislation being prepared	Under 16	Social media; broader "digital platforms" framing	Mandatory age verification; additional accountability measures signalled (Reuters)
United Kingdom	Consultation on possible under-16 restrictions and additional controls (AI chatbots, gaming, curfews) alongside existing *Online Safety Act 2023* duties	Consultation launched 2 Mar 2026 (closes 26 May 2026); existing safety regime in force	Under 16 (consultation topic)	Social media; potential extension to gaming platforms and AI chatbots	Consultation on bans/curfews; statutory platform duties enforced by Ofcom under OSA 2023 (GOV.UK)

Note: In Australia and the European proposals, "ban" typically functions as a platform obligation (to prevent underage accounts or access) rather than a criminal prohibition on minors themselves. Australia's government has emphasised that obligations and penalties fall on platforms, not on children or parents.

2.2. How Each Jurisdiction Is Approaching the Problem

2.2.1. Australia – Access restriction through platform liability

Australia is the only jurisdiction in this comparison to have enacted and commenced an access-restriction model, placing it at the frontier of ban-based approaches and making it the most instructive reference point for assessing practical enforceability.

The *Online Safety Amendment (Social Media Minimum Age) Act 2024* requires major platforms to take reasonable steps to prevent under-16 users from holding accounts, with potential civil penalties of up to A$49.5 million for non-compliance. The compliance burden falls squarely on platforms, not parents or children.[13]

A notable design feature is that the regime is operationally dynamic. Platform designation is governed by the *Online Safety (Age-Restricted Social Media Platforms) Rules 2025*, a subordinate instrument made by the Minister for Communications – meaning the list of covered platforms can evolve without requiring amendments to the primary legislation.[14] This has already generated public debate about whether services such as YouTube and other streaming platforms fall within the scope of the legislation,[15] and industry contestation over classification decisions.[16]

The broader governance challenge this approach illustrates recurs across jurisdictions: defining what counts as "social media" for the purpose of age-based access restrictions is inherently contentious.

Post-commencement, Australia's regulatory posture appears to be broadening from social media into adjacent AI-mediated services, including search engines and app stores. This signals a policy trajectory in which the "age restriction" lever may extend beyond classic social platforms into AI chat or content services.[17] Nonetheless, it is worth noting that the enforceability of restrictions is yet to be established, as there have been no actions or sanctions for non-compliance given the recency of the regulations.

2.2.2. China – Regulating the conditions of use, not the fact of presence

China's approach is structurally distinct. Rather than treating minors' presence on digital platforms as the problem, it accepts that presence, and instead seeks to regulate the conditions under which minors engage.

[13] Online Safety Amendment (Social Media Minimum Age) Act 2024 (Cth). Federal Register of Legislation, Australia. www.legislation.gov.au/C2024A00151/latest/text
[14] Online Safety (Age-Restricted Social Media Platforms) Rules 2025 (Cth). Federal Register of Legislation, Australia.
[15] Law Society Journal. (2025). YouTube's exemption reversed as Australia expands social media ban. https://lsj.com.au/articles/youtubes-exemption-reversed-as-australia-expands-social-media-ban
[16] Information Age (ACS). (2025). Social media giants fight YouTube's under-16s exemption. https://ia.acs.org.au/article/2025/social-media-giants-fight-youtubes-under-16s-exemption.html
[17] Reuters. (2026, 1 March). Australia says it may go after app stores, search engines, AI in age crackdown. www.reuters.com/business/media-telecom/australia-says-it-may-go-after-app-stores-search-engines-ai-age-crackdown-2026-03-01

In late 2025, China issued guidance on establishing "Minors Modes", which include default settings on usage times, durations, content, and functions, together with parental oversight features.[18] Additional measures have introduced classification rules for online information affecting minors' physical and mental health, with implementation extending into 2026.[19]

This creates a regulatory model that targets the conditions of use rather than access itself – and an enforcement surface that is not limited to a single platform class. The obligations apply across online products and services, including recommendation-driven content feeds, games, and other interactive services that minors commonly use.

China remains politically relevant to the "ban" debate: there are indications of proposals emerging in 2026, including a reported proposal during the "Two Sessions" to restrict social media use for under-16s.[20]

"Stricter regulatory approaches such as China's can strengthen control and reduce certain risks, but may simultaneously constrain children's participation and access to information."

– **Maryam Ehsani**, CEO, Child Safe Me

From a child-rights perspective, however, China's model raises its own tensions. Maryam Ehsani of Child Safe Me notes that stricter regulatory approaches can strengthen control and reduce certain risks, but may simultaneously constrain children's participation and access to information. This trade-off is not always made explicit in the policy rationale.[21]

2.2.3. France – When EU law meets national ambition

France offers the clearest illustration of how a national access-restriction model can be blocked at the implementation stage by supra-national legal constraints – and how policymakers attempt to learn from that experience.

The 2023 law, which established a "digital majority" principle requiring parental consent for social media use by under-15s, was never enforced in practice.[22] The implementing decree never took effect due to European Commission concerns about its compatibility with the country-of-origin principle under the *E-Commerce Directive*.[23] The 2026 bill is partly an attempt to remedy this enforcement gap on a more durable legal footing.

[18] China Law Translate. (2025). Minors modes. www.chinalawtranslate.com/en/minors-modes/
[19] Cyberspace Administration of China et al. (2025, 26 December). Measures for the classification of online information that may affect the physical and mental health of minors (effective 1 March 2026). English translation available at China Law Translate. www.chinalawtranslate.com/en/not-suitable-for-minors-final
[20] TechNode. (2026, 4 March). China adviser proposes banning social media use for under-16s during 2026 Two Sessions. https://technode.com/2026/03/04/china-adviser-proposes-banning-social-media-use-for-under-16s-during-2026-two-sessions
[21] See section Annex: Contributors.
[22] Le Monde. (2023, 29 June). France requires parental consent for under-15s on social media. www.lemonde.fr/en/france/article/2023/06/29/france-requires-parental-consent-for-under-15s-on-social-media_6039514_7.html
[23] Lexology. (2024). France: Digital majority and parental consent for minors on social media. www.lexology.com/library/detail.aspx?g=476cc5e0-0b83-4684-b335-ed2e9ed70589

In January 2026, France's National Assembly passed a bill banning social media for under-15s by a vote of 130 to 21.[24] The bill would require platforms to block access via age-verification mechanisms compliant with EU law, with proponents targeting enforcement from September 2026 (the start of the school year). The bill has since proceeded to the Senate.

France's experience highlights a distinctive feature of European policymaking: national measures must operate within EU frameworks on privacy, proportionality, and data protection – or risk being unenforceable.

2.2.4. Spain – Strong rhetoric, early-stage legislation

Spain's proposal is notable less for its legal maturity – the legislative text remains in preparation – than for its political framing. The government announced in early February 2026 that it would seek to prohibit social media for under-16s and require robust age verification, positioning these measures as part of a broader tightening of digital governance and content accountability.[25]

A distinguishing feature is the way in which access restrictions have been linked to broader debates about platform accountability and algorithmic amplification of harmful content.

As the legislative text develops, a key comparability question will be whether Spain's eventual legal instrument targets "social media" narrowly or extends into messaging, video-sharing, and interactive gaming.

2.2.5. United Kingdom – Layering new restrictions onto an existing duty-based regime

The UK case is distinctive because it already has a functioning statutory platform-duty regime – the *Online Safety Act 2023* (OSA) – which any new age-based restriction law would have to complement, creating a hybrid model with more enforcement infrastructure than most of its counterparts.[26]

Under the OSA, platforms are already required to conduct child safety risk assessments, implement risk-proportionate mitigations, and comply with enforceable codes of practice issued by communications regulator Ofcom, with powers to impose fines of up to £18 million or 10 per cent of global annual turnover.[27] Even without a ban, meaningful platform obligations targeting children are already in force and being actively enforced.

In March 2026, the UK government launched a public consultation considering whether social media should be banned for under-16s, and whether these same restrictions should also apply to gaming platforms and AI chatbots.[28] The scope is

[24] Reuters. (2026, 26 January). France's National Assembly debates banning under-15s from social media. www.reuters.com/sustainability/society-equity/frances-national-assembly-debates-banning-under-15s-social-media-2026-01-26/
[25] Reuters. (2026, 3 February). Spain to hold social media executives accountable for illegal, hateful content. www.reuters.com/world/spain-hold-social-media-executives-accountable-illegal-hateful-content-2026-02-03/
[26] Online Safety Act 2023, c. 50. UK Public General Acts. www.legislation.gov.uk/ukpga/2023/50/contents/enacted
[27] Online Safety Act 2023, c. 50. UK Public General Acts. www.legislation.gov.uk/ukpga/2023/50/contents/enacted
[28] UK Department for Science, Innovation and Technology. (2026, 2 March). Landmark consultation seeks views on major measures to protect children on social media, gaming platforms and AI chatbots. GOV.UK. www.gov.uk/government/news/landmark-

deliberately broad – reflecting recognition that the next generation of child safety risks may not sit neatly within legacy platform categories.

What is analytically significant is that the consultation outcome will determine whether an access-restriction layer is added on top of the existing duty-based architecture, or whether the government concludes that the OSA framework, properly enforced, is sufficient. This makes the UK a useful test case for a question that runs across all five jurisdictions: whether ban-based and duty-based approaches are genuinely complementary, or whether a well-resourced duty regime reduces the marginal value of an age-based access restriction.

Feature	Australia	China	France	Spain	United Kingdom
Core regulatory lever	Age access restriction (platform must prevent under-16 accounts)	Systems-based minors protection (modes, classification, platform duties)	Legislative move from parental consent model towards under-15 ban	Proposed under-16 ban with strong age assurance framing; broader accountability rhetoric	Consultation on under-16 ban/curfews alongside existing statutory platform duties
Legal status	Implemented (Dec 2025)	In force / staged measures effective 2026	Bill passed lower house; pending Senate	Announced; legislation to be introduced	Consultation launched (Mar 2026); existing OSA regime in force
Primary regulated entity	Social media platforms (designated)	Platforms, app stores/device ecosystem, content publishers	Social media / social networking services	Social media and potentially broader digital platforms	Online services within OSA scope; consultation covers wider categories
Enforcement architecture	Civil penalties; “reasonable steps” standard; no penalties on children/parents	Regulatory compliance; classification/notice requirements; structured minors-mode expectations	Age verification compliant with EU law; regulatory operationalisation required	Age verification; further accountability measures signalled	Ofcom enforcement of statutory duties; consultation explores bans/curfews
Treatment of AI	Post-ban pivot towards age assurance for AI services	AI governance integrated via classification/Minors Modes	Primarily social media-focused; AI addressed indirectly through EU ecosystem	Linked in public narrative to AI-generated harms including deepfakes	Consultation explicitly covers AI chatbots
Primary rationale	Child safety and youth wellbeing / mental health	Child protection within broader cyberspace governance; structured safety controls	Child protection and wellbeing; EU legal constraints	Child safety plus broader platform accountability	Child protection; consultation explores multiple levers beyond a ban

Note: “Age assurance” is used throughout this paper as an umbrella term covering age verification and age estimation approaches. Privacy-preserving architectures typically emphasise attribute-only proofs and minimised retention, including selective disclosure and zero-knowledge proof techniques.

consultation-seeks-views-on-major-measures-to-protect-children-on-social-media-gaming-platforms-and-ai-chatbots. Full consultation document available at www.gov.uk/government/consultations/growing-up-in-the-online-world-a-national-consultation.

2.3. Where the Five Jurisdictions Align

Convergences. Three shared patterns emerge. First, "platform safety for minors" is no longer treated as a purely voluntary corporate responsibility. Each jurisdiction re-allocates responsibility from parents and individual users to service providers, requiring platforms to change access, design, or operational controls. Australia's framework explicitly assigns the onus to platforms. The UK consultation similarly emphasises system-level interventions rather than user conduct.

Second, age assurance is gaining centrality. France's bill requires age verification compliant with EU law. Spain frames its proposal around "real barriers" through age verification. Australia's implementation relies on platforms identifying and removing underage accounts.

Third, social media bans are increasingly treated as components of broader digital safety packages. Spain links restrictions with wider platform accountability, the UK consultation extends to AI chatbots and gaming platforms, and Australia's regulatory focus is already expanding towards AI services.

Divergences. The most fundamental divergence is between a binary access restriction and a graded protection architecture. Australia, France, and Spain converge on a model where children below a threshold age should not access certain platforms. China builds a model that assumes minors will be online and modifies the environment through "minors mode" controls and content classification. The UK sits closer to the platform-duty end, with ban-like measures under consultation rather than enacted.

A second divergence concerns how regulatory obligations are distributed. Ban-based models concentrate compliance around age verification and onboarding controls. Systems-based models distribute obligations across moderation, recommendation systems, interface design, and operational governance. China operationalises this distribution through minors-mode guidance and classification measures. The UK channels compliance through risk assessments and mitigations across platform functions.

Third, legal interoperability constraints shape implementation. France's Bill is explicitly reported as needing EU-law-compliant age verification. Spain would face similar constraints. The practical consequence is that policy debates increasingly shift from the normative question – should minors be restricted? – towards the operational question – can age assurance be implemented accurately, privately, and resiliently?

2.4. Gaps, Blind Spots, and Side-Effects

2.4.1. The Age Assurance Dilemma: Civil Liberties, Cybersecurity, and Technical Feasibility

Across jurisdictions, age assurance has emerged as the central "hinge" issue shaping both public legitimacy and practical enforceability.

The underlying tension is political as much as technical. As Dr Chew Han Ei of the Institute of Policy Studies observes, there is a real risk that policymakers favour measures that look firm over measures that actually work: a ban is easy to announce, an age gate is easy to point to, and both create the appearance of grip. But depending on design and enforcement, they may also create fresh privacy risks or normalise verification systems far more intrusive than advocates will admit.[29]

"There is a real risk that policymakers favour measures that look firm over measures that actually work: a ban is easy to announce, an age gate is easy to point to, and both create the appearance of grip. But depending on design and enforcement, they may also create fresh privacy risks or normalise verification systems far more intrusive than advocates will admit."

– **Dr Chew Han Ei**, Head, Governance and Economy, Institute of Policy Studies

Civil liberties are a key focus of critics of the various models: if access to mainstream digital services requires presenting identity documents, biometrics, or other sensitive attributes, age-gating may constrain anonymous or pseudonymous participation – with implications for freedom of expression, association, and access to information.[30]

A second cluster of concerns relate to cybersecurity.[31] Age verification regimes can create new concentrations of sensitive data (identity documents, biometrics, device identifiers) that are attractive targets for breach. Critics point to the track record of data incidents and argue that expanding identity collection for routine platform access increases systemic exposure.[32]

Australia's regime has triggered technical debate about what platforms can realistically do to verify age without resorting to invasive methods, and whether the absence of proven privacy-preserving options may push providers towards more intrusive workarounds such as behavioural profiling or facial analysis.[33]

[29] See section Annex: Contributors.
[30] For an international human rights framing, see Article 19. (2024). Age assurance and human rights. www.article19.org/resources/age-assurance-human-rights/. For the US constitutional debate, see Electronic Frontier Foundation. (2025). The Supreme Court's decision on age verification tramples free speech and undermines privacy. www.eff.org/pages/supreme-courts-decision-age-verification-tramples-free-speech-and-undermines-privacy. Both argue that mandatory identity disclosure as a precondition for accessing mainstream digital services constrains anonymous and pseudonymous participation, with implications for freedom of expression, association, and access to information.
[31] The Conversation. (2025). Online age checking is creating a treasure trove of data for hackers. https://theconversation.com/online-age-checking-is-creating-a-treasure-trove-of-data-for-hackers-268586
[32] Electronic Frontier Foundation. (2024). Hack of age verification company shows privacy danger of social media laws. www.eff.org/deeplinks/2024/06/hack-age-verification-company-shows-privacy-danger-social-media-laws
[33] ASPI Strategist. (2025). Technical challenges in Australia's under-16s social media ban. www.aspistrategist.org.au/technical-challenges-in-australias-under-16s-social-media-ban

The Future of Privacy Forum (FPF) summarises common risk categories: excessive collection and retention, secondary use of identity data, false positives and negatives, interoperability gaps, breach risk, and uneven user acceptance – each capable of undermining either safety goals or rights protections if not explicitly governed (FPF, 2026).[34]

In response, a parallel strand of technical commentary argues that privacy risks are not inherent to age assurance but depend on implementation choices – particularly whether the system verifies age as an attribute ("over 16") rather than identity as a person. An IETF architecture draft has highlighted architectures using selective disclosure or zero-knowledge proofs (Knodel, 2024).[35] Recent policy analyses have emphasised deletion-after-verification and strict security controls as conditions for reducing privacy harm (Podda et al., 2024).[36]

Even where jurisdictions diverge on whether restrictions should exist, there is growing convergence around a principle: if age checks are mandated, they should be implemented using privacy-preserving patterns that minimise retained identifiable data and reduce 'linkability' across services.

2.4.2. Displacement, Definitional Arbitrage, and Cross-Border Enforcement

Even with robust age assurance, the scope and enforceability of platform restrictions depend on definitional clarity, ecosystem coverage, and effective enforcement across the digital service supply chain.

Implementation confirms that ban-like approaches create immediate pressures on verification systems.[37] Coverage of the French Bill has flagged enforceability challenges,[38] and Australia's early experience has generated significant discussion about circumvention[39] – with reports that many minors were able to retain or create accounts in the days following commencement.[40]

A recurring structural challenge is **definitional arbitrage**: where "social media" is the regulated category, services can contest classification, alter product features, or shift users to adjacent categories. Australia's evolving platform list illustrates that definitional scope is a governance question, not a neutral technicality.[41]

A related risk is **displacement**: restrictions on major platforms may push minors towards smaller, offshore, or less regulated services – potentially reducing the safety

[34] Future of Privacy Forum. (2026). Age assurance: Technologies and tradeoffs (updated infographic). https://fpf.org/blog/fpf-releases-updated-infographic-on-age-assurance-technologies-emerging-standards-and-risk-management/
[35] Knodel, M. (2024). Age architecture considerations. Internet Engineering Task Force, Internet-Draft. https://datatracker.ietf.org/doc/draft-knodel-age-arch
[36] Podda, E., Hölzmer, P., Amard, A., Sedlmeir, J., & Fridgen, G. (2024). The impact of zero-knowledge proofs on data minimisation compliance of digital identity wallets. *Internet Policy Review*. https://policyreview.info/articles/analysis/impact-zero-knowledge-proofs
[37] CNBC. (2025, 10 December). Australia bans social media for teens under 16 — here's what to know. www.cnbc.com/2025/12/10/australia-16-year-old-teens-ban-social-media-policy-law-ig-tiktok-fb-reddit-youtube-snapchat.html
[38] Euronews. (2026, 26 January). France could ban under-15s from social media within months — here's what we know. www.euronews.com/next/2026/01/26/france-could-ban-under-15s-from-social-media-within-months-heres-what-we-know
[39] NPR. (2025, 10 December). Social media ban for children in Australia. www.npr.org/2025/12/10/nx-s1-5639694/social-media-ban-children-australia
[40] PBS NewsHour. (2026). Social media platforms removed 4.7 million accounts after Australia banned them for children younger than 16. www.pbs.org/newshour/world/social-media-platforms-removed-4-7-million-accounts-after-australia-banned-them-for-children-younger-than-16
[41] eSafety Commissioner. (2026). Which social media platforms are age-restricted? Australian Government. www.esafety.gov.au/about-us/industry-regulation/social-media-age-restrictions/which-platforms-are-age-restricted

benefits of mature platforms' moderation infrastructure. This concern is often cited by child-safety advocates as a reason to prioritise safety-by-design obligations over blanket exclusions.[42]

A third challenge is **cross-border enforceability**. Because most major digital services operate globally, national restrictions rely on enforcement at local chokepoints – platform operators, app stores, identity providers, and payment intermediaries. This can create pressure to expand the regulatory perimeter when direct platform compliance proves difficult to verify.

2.4.3. The AI Exposure Gap Beyond Social Media

Even where social media access is restricted, minors may still encounter AI-generated harms through other channels, such as group chats, file-sharing platforms, gaming communities, image-generation tools, or embedded assistants.

Research with young people reinforces this point empirically. Dr Nikhila Natarajan's work in this area shows that children's media inventories typically span seven to twelve applications in constant flux, defying the neat platform categories that regulatory definitions assume.[43]

In this context, AI functions as a horizontal technology ("Ambient AI") threading across children's media landscape from social media to streaming services and generative AI tools. If the regulatory target is "social media" but the actual risk surface is a shifting constellation of AI-mediated services, the gap between policy scope and lived experience will widen with each new product cycle.

Australia's reported expansion towards AI services and age protection indicates how quickly policy attention can migrate from "social media" into a wider AI ecosystem. As Michael des Tombe of Netsafe observes, Australia's social media ban – despite being one of the most ambitious recent interventions in platform regulation – has a 'focus on restricting youth access to specified platforms [*that*] does not address harms occurring beyond those environments, such as those arising from AI chatbots and generative tools'.[44]

Meanwhile, the UK consultation does flag AI chatbots explicitly. China's approach, which applies across online products and services, is structurally better suited to this category-crossing problem – though it raises its own questions about breadth, compliance burden, and operational implementation.

[42] The Guardian. (2026, 1 March). UK teenagers to pilot social media ban and smartphone restrictions. www.theguardian.com/uk-news/2026/mar/01/uk-teenagers-pilot-social-media-ban-smartphone-restrictions
[43] See section Annex: Contributors.
[44] See section Annex: Contributors.

What This Means for ASEAN

The regulatory responses documented in Section 2 are not of merely comparative interest to ASEAN. They are part of the same global policy moment – and ASEAN is an active participant in it, not a latecomer.

The conditions that prompted regulatory action in Australia, France, Spain, and the United Kingdom – widespread youth social media use, evidence of algorithmic harm, the limits of self-regulatory models – are present across Southeast Asia too, often at greater demographic scale and with distinctive institutional contexts that call for regionally grounded responses.

Indeed, several ASEAN member states are already further along than some of the jurisdictions profiled in Section 2. Indonesia's regulation took effect on 28 March 2026, making it the first country in Southeast Asia to enforce a nationwide access restriction at this scale. Malaysia has enacted comprehensive online safety legislation. Singapore is implementing app-store-level age assurance obligations. Thailand, Vietnam, and the Philippines are each advancing national frameworks through their own legislative processes.

3.1. ASEAN's Youth Population and Digital Exposure

According to the ASEAN Secretariat, youth aged 15 to 34 constitute approximately 34 per cent of the region's population – around 213 million people, projected to peak at over 220 million by 2038 (ASEAN Secretariat, n.d.).[45] When children under 15 are included, the population directly implicated in child online safety debates is considerably larger.

This demographic scale coincides with some of the world's highest social media adoption rates: nearly 61.5 per cent of Southeast Asia's total population are active social media users – well above the global average (We Are Social & Meltwater, 2025).[46]

The exposure profile is correspondingly acute. Research by Radovanović (2026) indicates that children in Indonesia spend an average of 5.4 hours per day online, primarily on social platforms, with approximately 50 per cent of Indonesian children online having encountered sexual content on social media.[47] In Malaysia, police recorded RM2.7 billion in reported scam losses between January and November 2025, with significant youth victimisation (*Malay Mail*, 2025).[48]

These figures reflect structural conditions across the region, where platform penetration has outpaced both regulatory capacity and digital literacy infrastructure.

[45] ASEAN Secretariat. (n.d.). Education & youth. https://asean.org/our-communities/asean-socio-cultural-community/education-youth/
[46] We Are Social & Meltwater. (2025). Digital 2025: Southeast Asia. Reported in TechNode Global. https://technode.global/2025/02/11/digital-2025-nearly-two-thirds-of-southeast-asias-population-are-on-social-media
[47] Radovanović, D. (2026). Navigating child online protection in Indonesia: International norms, local realities, and the TikTok factor. *Global Social Welfare*, 5(6). https://link.springer.com/article/10.1007/s44206-025-00244-0
[48] Malay Mail. (2025, 8 December). Malaysians swindled out of RM2.7b in cyber scams in just 11 months, police data reveals. www.malaymail.com/news/malaysia/2025/12/08/malaysians-swindled-out-of-rm27b-in-cyber-scams-in-just-11-months-police-data-reveals/201156

Any regulatory model premised on age-based access restrictions will encounter a large, heterogeneous affected population with uneven access to formal identity documentation.

The ASEAN context is not a scaled-up version of the Australian or UK situation – the governance challenges are, in important respects, different in kind.

3.2. The Regional Policy Architecture

ASEAN's digital governance framework provides a strong foundation for child online safety action. The question is less whether that foundation exists than how it can be built upon with the specificity this policy area increasingly demands.

The *ASEAN Digital Masterplan 2030 (ADM 2030)*, adopted at the 6th ADGMIN in Hanoi in January 2026, sets ASEAN's strategic digital cooperation direction for 2026–2030, with trust, consumer protection, and platform accountability among its clearest areas of convergence.[49]

The *ASEAN Guidelines for Digital Platform Regulation*, welcomed at the same meeting, similarly emphasise platform accountability and consumer protection.[50] Neither instrument currently contains specific thresholds or obligations for the protection of minors – though the architecture they establish is well suited to accommodating such provisions.

The *ASEAN Digital Economy Framework Agreement (DEFA)*, still under negotiation, has potential to establish binding minimum obligations for child-safe platform design as part of its broader data governance provisions – an opportunity that negotiators may wish to explore as the agreement takes shape.[51]

The EU comparison offers a useful reference. The *Digital Services Act*'s guidelines on the protection of minors prescribe recommender system adaptations, restrictions on manipulative design, and age-appropriate interface requirements (European Commission, 2025).[52]

Developing equivalent operational specificity within ASEAN's own platform regulation guidelines – whether through amendment, through DEFA, or through a dedicated regional child safety instrument – is one option that merits consideration within the ADM 2030 implementation cycle.

The 2025 ASEAN ICT Forum on Child Online Protection, co-hosted by Malaysia and UNICEF in Kuala Lumpur, demonstrated what genuine multi-stakeholder design can look like: policymakers, platforms, civil society, and youth representatives co-designing standards in a single forum.[53] There is a strong case for making this kind of

[49] ASEAN Secretariat. (2026). ASEAN Digital Masterplan 2030. https://asean.org/book/asean-digital-masterplan-2030
[50] Bower Group Asia. (2026). ASEAN Digital Ministers Meeting 2026: Where regional ambition meets national priorities. https://bowergroupasia.com/asean-digital-ministers-meeting-2026-where-regional-ambition-meets-national-priorities/
[51] Nation Thailand. (2025). ASEAN Digital Economy Framework Agreement. www.nationthailand.com/business/tech/40057106
[52] European Commission. (2025). Guidelines on the protection of minors under Article 28 of the Digital Services Act. www.agcom.it/sites/default/files/media/allegato/2025/EU%20Guidelines_for_minors_online_28%20DSA.pdf
[53] UNICEF Malaysia. (2025, 18 November). ASEAN unites to build safer digital spaces for every child. www.unicef.org/malaysia/press-releases/asean-unites-build-safer-digital-spaces-every-child

inclusive engagement a regular feature of ASEAN's child safety governance, rather than a one-off initiative.

3.3. What ASEAN Member States Are Already Doing

While regional frameworks remain at the level of principles, several member states have moved quickly at the national level. The landscape as of March 2026 is summarised below.

Country	Age Threshold / Scope	Legal Status (Mar 2026)	Enforcement Mechanism	Primary Rationale
Indonesia	Under 16 (high-risk platforms); 13+ on lower-risk platforms	Government regulation signed Mar 2026; enforcement from 28 Mar 2026	Phased deactivation; Komdigi oversight; staged platform compliance	Cyberbullying, pornography, fraud, addiction – framed as "digital emergency"
Malaysia	Under 16	Online Safety Act 2025 in force (Jan 2026); eKYC via subsidiary legislation expected Q2 2026	MCMC oversight; mandatory eKYC (MyKad / national digital ID); Risk Reduction and Child Protection Codes in consultation	Child safety: cyberbullying, CSAM, scams; modelled closely on Australia
The Philippines	No ban enacted; Senate Bills 40 and 595 (2025) propose parental consent for minors' social media access	Under consideration	Proposed: verified parental consent; OSAEC and CSAEM Acts already in force for CSEA	Child sexual exploitation; content harm
Singapore	No blanket age ban; age assurance at app-store level for age-inappropriate apps	Code of Practice for Online Safety – App Distribution Services (effective 31 Mar 2025)	IMDA oversight; app stores must establish age with reasonable accuracy; annual safety reporting	Harm reduction and safety-by-design; "scaffolding" rather than exclusion
Thailand	Under 14 proposed as floor for personal accounts; usage caps under discussion	Ministry of Digital Economy and Society (DES) drafting rules (as of early 2026)	Under development	Mental health; online addiction; harmful content
Viet Nam	No age ban; content classification, identity verification, and cross-border accountability obligations	Decree 147/2024; Decision 88/QD-BTTTT (2025); 2025 Personal Data Protection Law	Ministry of Information and Communications; platform compliance obligations	Misinformation; content harm; online exploitation

Note: Brunei, Cambodia, Laos, Myanmar, and Timor-Leste are not profiled individually due to limited publicly available legislative data. All ASEAN Member States are party to the ASEAN Declaration on the Protection of Children from all Forms of Online Exploitation and Abuse. Timor-Leste, which has been accepted "in principle" for ASEAN membership, is not yet a full member of the organisation.

Three country approaches warrant closer attention, as they illustrate the range of regulatory models now being tested within the region.

Indonesia has enacted the most consequential measure. Its regulation introduces a risk-tiered model: children under 13 are excluded from all covered platforms, minors aged 13 to 15 may access lower-risk services, and all under-16s are barred from "high-risk" platforms including YouTube, TikTok, Instagram, and Facebook.[54] Platform classification as "high-risk" is based not only on content type but on behavioural architecture – specifically, whether the platform deploys engagement-maximising systems that can produce addiction or psychological harm.[55]

Civil society reactions have been divided. Amnesty International Indonesia described the measure as "overly simplistic" and flagged its tension with Indonesia's Convention on the Rights of the Child (CRC) obligations on children's rights to expression and access to information – a critique applicable to ban-based frameworks across the region (Amnesty International Indonesia, 2026).[56]

Malaysia's *Online Safety Act 2025* provides the legislative framework for an under-16 social media ban to be implemented through subsidiary legislation expected in Q2 2026.[57] Unlike Australia's "reasonable steps" standard, Malaysia's model leans towards mandatory electronic Know Your Customer (eKYC) mechanisms using government-issued identity documents, raising privacy concerns about data concentration and breach risk.[58]

The Act also includes a Risk Reduction Code and Child Protection Code – currently in consultation – which would impose structural obligations on algorithmic design and complaint mechanisms.[59]

Singapore has taken a different path. Its *Code of Practice for Online Safety – App Distribution Services*, effective March 2025, targets age-inappropriate content at the app-store level. App stores must establish user age with reasonable accuracy and restrict access to age-rated applications accordingly.[60]

This upstream chokepoint model is harder to circumvent than platform-level controls and avoids large-scale identity data collection – though it does not address harms within platforms that users are permitted to access. Singapore has signalled it is

[54] The Diplomat. (2026, March). Indonesia announces social media ban for children under 16. https://thediplomat.com/2026/03/indonesia-announces-social-media-ban-for-children-under-16/
[55] TechCrunch. (2026, 6 March). Indonesia outlines plan to limit under-16s' access to social media. https://techcrunch.com/2026/03/06/indonesia-outlines-plan-to-limit-under-16s-access-to-social-media/
[56] Amnesty International Indonesia. (2026). Blanket ban on social media for children: An overly simplistic response to online harms. www.amnesty.id/kabar-terbaru/siaran-pers/blanket-ban-on-social-media-for-children-an-overly-simplistic-response-to-online-harms/03/2026/
[57] ExpatGo. (2026, 26 February). Malaysia moves toward mandatory age verification for social media under Online Safety Act 2025. www.expatgo.com/2026/02/26/malaysia-moves-toward-mandatory-age-verification-for-social-media-under-online-safety-act-2025/
[58] South China Morning Post. (2026). Malaysia's bid to bar under-16s from social media using ID checks stokes privacy fears. www.scmp.com/week-asia/lifestyle-culture/article/3333904/malaysias-bid-to-bar-under-16s-from-social-media-using-id-checks-stokes-privacy-fears
[59] Malaysian Communications and Multimedia Commission. (2026, 12 February). Public consultation on the codes of the duties under the Online Safety Act 2025: Draft Risk Mitigation Code and Child Protection Code. https://mcmc.gov.my/en/onsa/media/announcements/public-consultation-on-the-codes-of-the-duties-und
[60] Mayer Brown. (2025, December). Malaysia's proposed social media ban for children: How it compares with Australia and Singapore. www.mayerbrown.com/en/insights/publications/2025/12/malaysias-proposed-social-media-ban-for-children-how-it-compares-with-australia-and-singapore

studying bans similar to Australia's, suggesting the current approach may be a first phase rather than a final position.[61]

Singapore's phased posture reflects a broader scepticism, articulated domestically, about the efficacy of high-visibility measures. Dr Chew Han Ei cautions that protecting children online is not a question of doing more, loudly, but of doing the right things, with enough humility to recognise that some highly visible solutions may turn out to be poor ones.[62]

"Protecting children online is not a question of doing more, loudly, but of doing the right things, with enough humility to recognise that some highly visible solutions may turn out to be poor ones."

– **Dr Chew Han Ei**, Head, Governance and Economy, Institute of Policy Studies

3.4. The Case for Harmonisation

The divergence in national approaches raises an immediate question about regional coherence. If minors in one member state are excluded from major platforms while their counterparts in a neighbouring country are not, the regulatory benefit of the more restrictive framework is partially eroded. Platforms can contest classification, and users can migrate to services accessible through less restrictive regimes.

The case for some degree of regional harmonisation is therefore pragmatic rather than principled. A minimum common standard, such as the EU's *General Data Protection Regulation (GDPR's)* baseline digital age of consent, within which member states retain discretion to go further, would reduce arbitrage opportunities, strengthen the collective negotiating position of ASEAN regulators *vis-à-vis* global platforms, and give lower-capacity member states a regional mandate for action they might otherwise lack.

Whether that standard should take the form of a minimum age floor, a set of platform design obligations, or both is the deeper question. The table below applies the comparative frame from Section 2 to ASEAN's specific conditions.

[61] Ministry of Digital Development and Information, Singapore. (2025, 7 January). Response to Parliamentary Questions on social media platforms to ensure age-appropriate use for young children. www.mddi.gov.sg/newsroom/pqs-on-social-media-platforms-ensure-age-appropriate-use-young-children; see also Bloomberg. (2025, 7 January). Singapore engages Australia to study social media age limits. www.bloomberg.com/news/articles/2025-01-07/singapore-engages-australia-to-study-social-media-age-limits

[62] See section Annex: Contributors.

Consideration	Access Restriction / Ban Model	Regulatory / Platform Duty Model
Clarity of obligation	High: a clear age floor reduces ambiguity for platforms and parents	Variable: effectiveness depends on the quality of design obligations; harder to communicate publicly
Enforceability in ASEAN	Challenging: requires age assurance infrastructure; uneven digital ID penetration; circumvention likely without robust enforcement	More tractable: obligations can be scaled to platform risk and national enforcement capacity
Privacy and identity risk	Higher: ban models tend to require large-scale identity collection	Lower if well-designed: privacy-preserving defaults can be specified without mandatory identity verification
Child rights compatibility	Risk of over-exclusion: minors lose access to platforms carrying educational and social value	Preserves access while modifying conditions of use; more consistent with CRC principles
Protection against algorithmic harms	Indirect: removes access but does not address harms on adjacent or unregulated platforms	More direct: targets the design features (recommendation, engagement optimisation) that produce harms
Fit with ASEAN diversity	Poor fit for lower-capacity member states; enforcement gaps create patchwork compliance	Better calibrated: minimum standards can be agreed regionally while implementation varies by capacity
Regional harmonisation potential	Possible if thresholds are aligned, but definitional disputes persist	Stronger: platform duties can be set regionally with national flexibility, closer to the EU DSA model
Clarity of obligation	High: a clear age floor reduces ambiguity for platforms and parents	Variable: effectiveness depends on the quality of design obligations; harder to communicate publicly
Enforceability in ASEAN	Challenging: requires age assurance infrastructure; uneven digital ID penetration; circumvention likely without robust enforcement	More tractable: obligations can be scaled to platform risk and national enforcement capacity

The comparative evidence from Section 2 suggests that there is no single “right” model. The most effective approach in any given context depends on a range of factors: the maturity of digital identity infrastructure, the capacity of regulators to enforce obligations at scale, the depth of existing platform governance frameworks, and the broader institutional and cultural context.

In member states with well-established digital identity systems and regulatory infrastructure – such as Singapore, Malaysia, and potentially Thailand – a hybrid model combining an age floor with platform duty obligations appears both feasible and likely to produce durable protection.

In contexts where digital ID penetration is more limited or where enforcement resources are more tightly constrained, platform duty obligations focused on safety-by-design may deliver greater real-world child protection value than a ban that proves difficult to enforce consistently. These are not judgements about relative capacity – they are reflections of the reality that effective governance design must be calibrated to the conditions in which it operates.

In all cases, the displacement risk is structurally more acute in a multi-jurisdictional region than in single-country models. If restrictions in one member state have no equivalent in a neighbouring country, minors may simply shift to platforms accessible through less restrictive regimes. This is one of the strongest practical arguments for some form of regional coordination on platform design obligations – even where member states retain full discretion on access restrictions.

3.5. Multilateral Cooperation and Enforcement

Effective child online safety governance in ASEAN faces a structural enforcement challenge: the major platforms are not ASEAN companies. Compliance depends on commercial incentives or credible enforcement threats, and the latter require institutional mechanisms that most ASEAN member states do not yet possess individually.

Several pathways merit exploration. The ASEAN Telecommunications Regulators Council (ATRC) already coordinates platform regulation among national Information and Telecommunications (ICT) regulators; an expanded mandate encompassing joint child safety enforcement – shared evidence standards, coordinated investigations – could give it considerably more practical utility.[63]

The *Regional Plan of Action for the Protection of Children from All Forms of Online Exploitation and Abuse in ASEAN (ASEAN RPA on COEA)*, under development through the ICT Forum process, would benefit from the inclusion of specific and measurable platform accountability benchmarks alongside its aspirational language.[64]

Beyond the region, the Global Online Safety Regulators Network – comprising Ofcom, the eSafety Commissioner, and others – is an emerging forum for shared enforcement standards.[65]

Formal participation by ASEAN member state regulators would strengthen both the Network's global legitimacy and the region's access to enforcement best practices. The *UN Convention on Cybercrime*, signed by several ASEAN member states, provides mutual legal assistance relevant to cross-border child harm cases.[66] These external anchors are valuable precisely because they constrain the ability of global platforms to treat any single region as a lower-scrutiny market.

[63] ASEAN Secretariat. (n.d.). ASEAN digital sector. https://asean.org/our-communities/economic-community/asean-digital-sector
[64] UNICEF Malaysia. (2025, 18 November). ASEAN unites to build safer digital spaces for every child. www.unicef.org/malaysia/press-releases/asean-unites-build-safer-digital-spaces-every-child
[65] NBC News. (2026, 16 February). UK's Starmer seeks greater powers to regulate online access. www.nbcnews.com/world/united-kingdom/uks-starmer-seeks-greater-powers-regulate-online-access-rcna259203
[66] Krisanaraj, J. (2025, 21 October). ASEAN and UNESCO urge digital governance to tackle online threats. *Nation Thailand*. www.nationthailand.com/business/tech/40057106

3.6. Key Stakeholders

Effective governance requires coordinated engagement across actors whose interests and mandates do not naturally align. The matrix below maps the main stakeholders.

Stakeholder Group	Key Actors in ASEAN	Role	Key Tensions
Governments and Ministries	Communications/digital ministries; ICT regulators (MCMC, IMDA, Komdigi); Ministries of Education and Women/Family Affairs	Primary legislators and regulators	Aligning child safety with digital economy priorities; avoiding over-reach or under-enforcement
ASEAN Secretariat and Bodies	Working Group on AI Governance; ASEAN AI Safety Network Secretariat (KL); Senior Officials Meeting on Youth (SOMY)	Regional coordination and norm-setting	Navigating consensus among member states with divergent regulatory capacities
Technology Platforms	Meta, ByteDance (TikTok), Google (YouTube), Snap, X; regional platforms (Bigo Live, Lemon8)	Primary regulated entities	Compliance cost; definitional disputes; differential enforcement risk
Civil Society and Child Protection	UNICEF, Save the Children, Internet Watch Foundation; national child rights NGOs	Advocacy, monitoring, safeguarding	Balancing child safety with rights to expression and participation
Institutes of Higher Learning	NUS, Universiti Malaya, Chulalongkorn University; AIAPI; Tech For Good Institute	Evidence generation; policy evaluation; capacity building	Contextualising research to ASEAN conditions rather than extrapolating from Western studies
Statutory Boards and Regulators	IMDA (SG), MCMC (MY), Komdigi (ID); national data protection authorities	Operationalise legislation; enforce standards	Building technical capacity; managing cross-border jurisdiction
Parents, Carers, and Youth	Parent-teacher associations; youth advisory councils; children's consultative groups	Inform policy design; demand-side accountability	Ensuring youth voices shape policy rather than merely legitimise decisions already taken
Industry Associations and Standards Bodies	CCIA Asia; ISO/IEC (age assurance, ISO/IEC 27566-1); ITU	Develop interoperable standards; industry input on feasibility	Managing divergence between regional standards and global platform architectures

Two observations follow. First, the Tech For Good Institute's 2025 policy brief identifies the absence of cross-sector coordination as a primary obstacle to effective governance: communications ministries regulate platforms, education ministries address literacy, and social welfare ministries handle victim support – without the integration a comprehensive framework requires.[67]

Second, children and young people remain chronically under-represented in policy design. The 2025 ASEAN ICT Forum's decision to co-design its agenda with UNICEF's Young People's Advisory Group is a positive precedent – one that could usefully become a standing feature of ASEAN's child safety governance processes.

There is also a question of substance, not only process. Dr Nikhila Natarajan reports that dozens of teens in her research have been emphatic that AI-powered image manipulation is "creepy" and "wrong"; a clear signal that young people hold strong and articulate views on AI practices that directly affect them, and that those views can inform regulatory design if mechanisms exist to capture them.[68]

3.7. The ASEAN AI Safety Network

The establishment of the ASEAN AI Safety Network (ASEAN AI Safe), with its Secretariat based in Kuala Lumpur, is the most significant new institutional development in ASEAN's AI governance landscape.

The Network's mandate – AI safety research, best practices, capacity building, and multistakeholder collaboration – was designed as a broad AI governance mechanism. But the alignment with child online safety objectives is natural and practically important.[69]

> "Teens in my research have been emphatic that AI-powered image manipulation is 'creepy' and 'wrong'."
>
> – **Dr Nikhila Natarajan**, Adjunct Professor, School of Communication and Information, Rutgers, The State University of New Jersey

The harms most acutely affecting minors in digital spaces – algorithmic recommendation of harmful content, AI-generated child sexual abuse material, deepfake-mediated harassment, and AI companions operating without child-safe defaults – are fundamentally AI governance problems, not merely platform moderation failures. They cannot be adequately addressed through age verification alone; they require the systems-level safety assessment that ASEAN AI Safe is positioned to facilitate.

The Network is explicitly guided by the *ASEAN Responsible AI Roadmap (2025-2030)* and the *ASEAN Guide on AI Governance and Ethics* – frameworks containing

[67] Tech for Good Institute. (2025). Safeguarding the digital generation: Strengthening child online protection frameworks in Southeast Asia. https://techforgoodinstitute.org/wp-content/uploads/2025/12/Policy-Brief-Safeguarding-the-Digital-Generation_-Strengthening-Child-Online-Protection-Frameworks-in-Southeast-Asia.pdf
[68] See section Annex: Contributors.
[69] CybersecurityAsia. (2025). Malaysia champions ASEAN AI Safety Network. https://cybersecurityasia.net/malaysia-champions-asean-ai-safety-network

accountability and human rights provisions directly applicable to child safety.[70] Operationalising these in a child-specific context could give them practical traction they do not yet fully possess.

One avenue worth exploring: the ASEAN AI Safe Secretariat could establish a dedicated work stream on child and adolescent online safety, in partnership with UNICEF, the ASEAN ICT Forum child protection mechanism, and relevant national regulators. Such a work stream might develop a regional taxonomy of AI-mediated harms to minors, explore minimum standards for child-safe AI deployment across ASEAN markets, and pilot a shared monitoring mechanism aggregating harm data across member states.

This kind of initiative would also help address an important capacity question that several member states face: the technical expertise required to audit AI systems deployed by major platforms for child safety compliance is specialised and resource-intensive. Building that capability collectively – rather than expecting each member state to develop it independently – is one of the strongest arguments for a regional approach.

3.8. Four Questions Worth Addressing Together

The thread running through this section is that ASEAN is not at the beginning of this policy conversation – it is mid-stream, with two member states already implementing frameworks, others in active legislative development, and a regional AI governance architecture that provides a usable institutional foundation. The pace of national action is itself a sign of momentum and seriousness.

That momentum also raises a question: whether the region's collective interests are best served by continued parallel national experimentation, or whether there is value in deliberate coordination on at least some dimensions. Four strategic questions stand out as candidates for shared reflection.

First, **the architecture question**: whether child online safety is best approached primarily as an access restriction, a platform design obligation, or a hybrid of the two. The jurisdictions in Section 2 are testing all three approaches; ASEAN member states are now doing the same. A shared analytical framework for evaluating which approaches work under what conditions could help each member state make better-informed choices – without requiring uniformity.

Second, **the harmonisation question**: whether there is appetite for common minimum standards on age thresholds or platform obligations – and if so, through which instrument. Without some coordination, there is a risk that regulatory fragmentation creates arbitrage opportunities for platforms and undermines the effectiveness of even well-designed national frameworks.

Third, **the age assurance question**: both ban-based and duty-based models increasingly depend on reliable age determination. A regional approach to age assurance – one that is privacy-preserving, interoperable across member states, and

[70] GovInsider. (2025). The ASEAN AI Safety Network to advance AI safety for the region. https://govinsider.asia/intl-en/article/the-asean-ai-safety-network-to-advance-ai-safety-for-the-region

accessible to users without formal identity documents – could reduce duplication and strengthen each national framework's credibility. This is an infrastructure challenge that lends itself to collective investment, potentially aligned with the emerging *ISO/IEC 27566-1* standard.[71]

Fourth, **the institutional question**: which body is best placed to coordinate the child online safety agenda across ADM 2030, DEFA, ASEAN AI Safe, the ICT Forum, and national regulators. At present, this coordination is distributed across multiple bodies. One option worth considering is designating the ASEAN AI Safe Secretariat as the primary institutional home for the region's AI-mediated child safety work, with formal links to UNICEF, Senior Officials Meeting on Youth (SOMY), and the platform regulation bodies. Whether this is the right institutional design is itself a question best resolved through consultation with the relevant stakeholders.

[71] ISO/IEC 27566-1:2025 is the first international standard for age assurance systems, providing a framework for verifying or estimating user age responsibly. Published in December 2025, it covers terminology, stakeholder roles, and privacy-by-design.

Conclusions and Next Steps

This paper's comparative review points to a major shift in how governments approach children's digital governance. Policymakers are no longer relying primarily on voluntary platform controls. Instead, they are experimenting with bans, "minor modes", risk-based duties, and restrictions on manipulative AI practices.

The most urgent policy gap is not the existence of rules – it is the lack of robust, comparable evidence on which interventions measurably reduce harms to minors without generating unacceptable unintended consequences, especially for privacy, access, and inclusion.

The practical implication is that ASEAN's most valuable near-term contribution may lie in building regional evaluation capacity and coordinating minimum standards – aligned with ADM 2030, DEFA, and ASEAN AI Safe – rather than waiting for full legal harmonisation, which is neither realistic nor necessary.

4.1. Building an ASEAN Evidence Base on Online Safety

Across jurisdictions, legislative and regulatory activity is accelerating, but evaluation infrastructure remains weak.

The EU combines platform duties under the *Digital Services Act* with AI-specific constraints under the *AI Act*. The DSA Guidelines frame "safety and privacy by design" as an expectation for platforms accessible to children, guiding compliance assessment under Article 28(1).

The *AI Act* prohibits practices especially salient for minors, including AI systems that exploit age-related vulnerabilities and systems using purposefully manipulative or deceptive techniques that materially distort behaviour in ways likely to cause significant harm.

China has taken a more administrative approach, including algorithmic recommendation provisions requiring providers to develop "minor-appropriate" modes and to avoid using recommendations to induce internet addiction. Its generative AI measures require providers to prevent minor users' over-reliance on or addiction to generative AI services. Deep synthesis rules also impose labelling and technical obligations for synthetic content – relevant to deepfake-enabled abuse.

These approaches illustrate an expanding policy menu: bans or age gates; child-specific modes; recommender-system constraints; manipulative-design restrictions; content labelling; and platform risk management.

Yet policy debates still frequently rely on proxy indicators – headline compliance, platform reporting volumes, or single-country controversies – rather than comparable evidence on outcomes. The AIAPI–Netsafe paper makes an important foundational point: AI amplifies both the scale and personalisation of existing harms, which complicates governance because interventions targeting one surface (content

removal) can leave other drivers (algorithmic amplification, synthetic production) intact.

If harms are system-level and adaptive, policy without evaluation risks becoming symbolic (high salience, low impact) or counterproductive (shifting harms elsewhere, increasing surveillance burdens). Without evidence, governments face two symmetrical risks: under-enforcement that leaves harms unchanged, and over-reach that produces privacy and exclusion costs disproportionate to the benefits.

Critically, child online safety measures must be judged against child rights principles. The UN Committee on the Rights of the Child's *General Comment No. 25* affirms that children's rights must be respected, protected, and fulfilled in the digital environment, creating a clear evaluative standard: interventions should be proportionate, non-discriminatory, and attentive to children's participation and best interests (UN Committee on the Rights of the Child, 2021).[72]

4.2. A Research Agenda on What Actually Works

One of the clearest lessons from the comparative analysis is that regulation without evaluation risks producing outcomes that are either symbolic or counterproductive.

A comparative research agenda that is designed to answer three questions ('what works?', 'for whom?', and 'at what cost?') could help ASEAN avoid the pitfalls that are already becoming visible in other jurisdictions.

- **Mental health and well-being outcomes.** The political justification for bans is often linked to youth well-being. But mental health outcomes are multi-factorial and context-dependent, making simple pre-post comparisons unreliable. A common outcomes framework with supporting quasi-experimental evaluation where feasible (e.g. phased implementation across jurisdictions or age cohorts) alongside mixed-methods research capturing lived experience would give policymakers across the region a stronger evidence base.

  Such evaluation would benefit from measuring exposure to specific risk vectors (cyberbullying, sexual coercion, deepfake harassment), sleep disruption, social connectedness, help-seeking behaviour, and school engagement instead of relying on "screen time" as a proxy.

- **Behavioural shifts and displacement.** When access is restricted on major platforms, minors may migrate to less moderated spaces such as private messaging groups, niche forums, and grey-market apps. Tighter recommender constraints may reduce harmful exposure on feeds but increase search-based or peer-to-peer sharing if the underlying demand for social media and AI products remains.

[72] UN Committee on the Rights of the Child. (2021). General Comment No. 25 on children's rights in relation to the digital environment (CRC/C/GC/25). www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation

A credible evaluation programme should incorporate multi-platform behaviour mapping with safeguards against invasive surveillance. These will include privacy-preserving analytics, voluntary longitudinal panels, and partnerships with schools and youth services.

- **Variation across regulatory contexts.** The same intervention can perform differently depending on enforcement credibility, trust in government, digital identity infrastructure maturity, and the availability of child support services. Strong age assurance duties may be feasible where identity infrastructure is reliable and privacy oversight is credible, but exclusionary where populations lack formal identity documents or where marginalised youth depend on anonymity for safety.

  ASEAN's diversity is an analytic advantage. It enables comparative learning about which combinations of "ban" and "duty" levers reduce harm under different conditions. Across all strands, ethical constraints are paramount. Research must minimise data collection, protect confidentiality, avoid re-traumatisation, and include child-appropriate consent and safeguarding protocols. Measuring safety should not require creating new surveillance harms.

The harder questions, as Dr Chew Han Ei puts it, are less politically attractive: 'can the measure be implemented sensibly?', 'what new monitoring burdens does it create?', and 'what is the fallback if it fails?'; these are precisely the questions a regional research agenda should be designed to answer.[73]

4.3. Towards Minimum Standards Alignment

Full legal harmonisation on age thresholds or platform categories across ASEAN is neither likely nor, arguably, necessary in the near term.

What is achievable – and what the comparative analysis suggests would be valuable – is coordination on common minimum standards that reduce fragmentation, support cross-border enforcement, and raise baseline protections for minors across the region.

This approach is consistent with ASEAN's broader ambitions. ADM 2030 sets a shared strategic direction for 2026-2030. DEFA is framed as accelerating ASEAN's transformation into a leading digital economy. Online safety for minors can be positioned not as a brake on digital growth but as the trust infrastructure that makes digital integration politically sustainable.

Drawing on the comparative evidence and the policy considerations discussed throughout this paper, a practical minimum-standards package might rest on four pillars.

1. **Platform responsibility** could require large platforms to conduct child-specific risk assessments, implement child-appropriate defaults, provide accessible reporting and redress, and enable independent scrutiny of recommender

[73] See section Annex: Contributors.

impacts on minors. The EU's *DSA Guidelines* offer one model for how a rights-aware safety-by-design expectation can be articulated without requiring identical national laws – though ASEAN would naturally adapt any such approach to its own regulatory traditions and institutional context.

This logic aligns with what Michael des Tombe (Netsafe) describes as 'a duty of care framework requiring providers to identify, assess, and mitigate risks to children across their entire service offering' – an approach that is technology-agnostic by design and therefore more resilient to the pace of AI development than platform-category-based restrictions.[74]

> "A duty of care framework requiring providers to identify, assess, and mitigate risks to children across their entire service offering, coupled with greater international alignment, warrants serious consideration and is more likely to deliver sustained improvements in child safety online."
>
> – **Michael des Tombe**, Chief Legal Advisor, Netsafe

2. **Age assurance** could be guided by regionally agreed principles of proportionality, privacy-preservation, accessibility, and interoperability – with compatible approaches and mutual recognition pathways that prevent fragmented, high-risk implementations. The aim would be not to impose a single technical solution but to establish shared principles that give each member state room to implement in ways suited to its own conditions.
3. **Child-safe AI deployment** could establish expectations for systems likely to affect minors, including recommender systems and generative AI tools. The EU AI Act's prohibition on exploiting age-related vulnerabilities in Article 5[75] and China's explicit anti-addiction duties for algorithmic recommendations both provide useful comparative reference points – though the design of any ASEAN-specific approach would need to reflect the region's own balance between rights protection, innovation, and enforcement capacity.
4. **Cross-border enforcement** could focus on practical cooperation mechanisms: shared incident taxonomies, regulator-to-regulator information exchange, coherent evidence standards, and joint engagement with global platforms. Regulatory fragmentation is itself a safety risk – it incentivises arbitrage and weakens the credibility of enforcement across the region.

4.4. Proposed Directions for Discussion

The establishment of ASEAN AI Safe creates an institutional opportunity to host a dedicated child online safety and AI harms work stream, and the comparative evidence assembled here is offered as a contribution to that process.

The ASEAN AI Safe declaration explicitly recognises uneven institutional maturity across member states and calls for strengthened coordination among government, industry, academia, civil society, and communities; precisely the configuration that evidence-led child safety governance requires.

[74] See section Annex: Contributors.
[75] European Union. (2024). Article 5: Prohibited AI practices. In Artificial Intelligence Act. https://artificialintelligenceact.eu/article/5/

With that spirit in mind, we offer five possible directions for consideration and discussion:

(i) A standing multistakeholder forum under a suitable regional anchor (potentially ASEAN AI Safe), with explicit child-rights safeguards and meaningful youth participation – building on the model piloted at the 2025 ASEAN ICT Forum.

(ii) A minimum common measurement framework for child online safety and AI-mediated harms, enabling comparison across member states and providing a shared evidence base for policy evaluation.

(iii) Structured data access arrangements with major platforms for approved, privacy-preserving research conducted under independent ethics oversight – recognising that without platform data, evaluation of policy effectiveness will remain speculative.

(iv) An ASEAN minimum standards playbook – developed collaboratively – covering platform responsibility, age assurance principles, child-safe AI deployment, and cross-border cooperation. Such a document would function as a reference rather than a mandate, offering member states a shared vocabulary and analytical framework.

(v) A commitment to iterative review: treating interventions as policies to be tested, refined, and – where evidence warrants – adjusted or rolled back if harms or inequities outweigh benefits.

In practical terms, a programme of this kind could be launched relatively quickly: by agreeing on a small set of shared outcome metrics and ethical protocols, convening an inaugural regulator-platform-civil society-youth roundtable to set priorities and secure participation commitments, and commissioning cross-country pilot evaluations on specific policy levers (age assurance approaches, child-safe recommender defaults, deepfake response workflows).

An ASEAN baseline report mapping harms, interventions, and early results could be published within 6-9 months, creating a reference point for ADM 2030 and DEFA implementation discussions.

These are starting points, not conclusions. We welcome the opportunity to discuss and refine them further with ASEAN policymakers, regulators, and stakeholders.

Annex: Contributors

The table below captures the detailed perspectives offered by each of our contributing specialists:

Contributor	Contribution
Dr Nikhila Natarajan, Adjunct Professor, School of Communication and Information, Rutgers, The State University of New Jersey	Technology design must be downstream from a deep understanding of youth development. For nearly a century – from the Payne Fund Studies in the 1930s to today – concerns about new technologies' effects on youth have loomed large in both research and public debate. When public sentiment reaches a tipping point, regulation tends to focus on what technology is doing to children, rather than on children's developmental needs and perspectives. The central opportunity is to shift this "north star" by placing youth developmental characteristics and youth voices at the heart of technology design and policymaking. Past legislative efforts – the V-chip, COPPA – illustrate the limitations of "tech-first" approaches targeting specific content or data collection without addressing broader industry structures. When policymakers say they have banned social media, they overlook that children use an eclectic collection of media appealing to their peer sociality. Research with young people shows that their media inventory defies easy categorisation (7 to 12 apps, constantly shifting). AI is a horizontal technology – "Ambient AI" – threading across children's media landscape from social media to streaming and generative AI. Children's developmental needs and voices should drive technology design. Dozens of teens in my research have been emphatic that AI-powered image manipulation is "creepy" and "wrong". Technology design should be personalised to developmental needs, and solutions should be aligned to helping young people thrive.
Maryam Ehsani, CEO, Child Safe Me	From a child rights perspective, current regulations should be judged by how well they ensure safeguarding and enable meaningful participation – yet most fall short. Measures such as social media bans in the UK, Australia, and France may deliver short-term results but are unlikely to be effective without sustained investment in education and awareness. Drawing on 17 years of experience, we still struggle to embed preventive mechanisms in online child safety; responses remain reactive. Stricter approaches such as China's strengthen control and risk reduction but raise concerns around participation and access to information. Responsible AI must be a core priority, not an add-on. Effective regulation should move towards child-centred, risk-based frameworks prioritising prevention, safety-by-design, accountability, and children's active involvement in decision-making. There is a clear need to move beyond restriction-based models and towards empowerment.
Dr Chew Han Ei, Head, Governance and Economy, Institute of Policy Studies	Few would disagree that protecting minors online is necessary. But there is a real risk that policymakers favour measures that look firm over measures that actually work. A ban is easy to announce; an age gate is easy to point to. Both create the appearance of grip. Some may reduce exposure to harm. But depending on design and enforcement, they may also create fresh privacy risks or normalise verification systems far more intrusive than advocates will admit. The harder questions are less politically attractive: Can the measure be implemented sensibly? What new data or monitoring burdens does it create? What is the fallback if it fails? Protecting children online is not just a question of doing more, loudly. It is a question of doing the right things, with enough humility to recognise that some highly visible solutions may turn out to be poor ones.
Amrin Amin, Head, Corporate Development, Temasek Foundation (Former Senior Parliamentary Secretary, Ministry of Health)	AI and social media can pose real risks to minors' attention, identity, and mental health. Rules and platform controls can help, but they are often blunt and can be bypassed – especially by increasingly tech-savvy minors. A holistic perspective is important: combining thoughtful platform design, transparency, and accountability with guidance from parents, schools, and communities. Equally crucial is giving young people space to build resilience and social skills – to play in nature, talk face-to-face, and experience life beyond screens.
Michael des Tombe, Chief Legal Advisor, Netsafe	As jurisdictions around the world consider how to regulate artificial intelligence for child safety, a familiar problem re-emerges: digital harms are transnational, while regulation remains domestic, fragmented, and reactive. This challenge is further compounded by the pace of generative AI development, meaning regulatory interventions risk becoming outdated on arrival. Australia's social media ban is a case in point. Despite being one of the most ambitious recent interventions in platform regulation, its focus on restricting youth access to specified platforms does not address harms occurring beyond those environments, such as those arising from AI chatbots and generative tools. While Australia's approach provides a useful policy test case, effective, enduring, and technology-agnostic regulation requires something different. A duty of care framework requiring providers to identify, assess, and mitigate risks to children across their entire service offering, coupled with greater international alignment, warrants serious consideration and is more likely to deliver sustained improvements in child safety online

Access Partnership Institute

AI ASIA PACIFIC INSTITUTE

Access Partnership offices

Europe

London

The Tower, Buckingham Green
Buckingham Gate
London, SW1E 6AS
United Kingdom

+44 20 3143 4900
london@accesspartnership.com

Brussels

8th Floor, Silversquare Europe
Square de Meeûs 35
Bruxelles, 1000
Belgium

brussels@accesspartnership.com

North America

Washington DC

1300 Connecticut Avenue NW
Suite 250
Washington, DC 20036
USA

+1 202 503 1570
washingtondc@accesspartnership.com

Asia

Singapore

Asia Square, Tower 2
#11-20
12 Marina View
Singapore 018961

+65 8323 7855
singapore@accesspartnership.com

Jakarta

Revenue Tower 21st Floor
Unit 104 SCBD Lot 13,
Jl. Jend. Sudirman Kav. 52-53
Provinsi DKI Jakarta, 12190
Jakarta, Indonesia

+62 21 5020 0949

Kuala Lumpur

Common Ground Q Sentral
Level 39, Unit 39-02 (East Wing), 2A,
Jalan Stesen Sentral 2, Kuala Lumpur
Sentral, 50470
Kuala Lumpur, Malaysia

Bangkok

188 Spring Tower
11th Floor, Unit 106, Phayathai
Road, Thung Phayathai,
Ratchathewi, 10400
Bangkok, Thailand

+ 66 (2)-8216148

Hanoi

19th floor, Tower 1
Capital Place Building
No 29 Lieu Giai Street
Ngoc Khanh Ward, Ba Dinh Diítrict
Hanoi, Vietnam

Manila

CG8ROCKWELL Level 21, 8
Rockwell, Hidalgo Dr., Rockwell
Center, Bgy. Poblacion, 1210
Makati City, Metro Manila
Philippines

Middle East and Africa

Abu Dhabi

Al Wahda City Tower, 20th Floor
Hazaa Bin Zayed The First Street
PO Box 127432
Abu Dhabi, UAE

abudhabi@accesspartnership.com

Johannesburg

119 Witch-Hazel Avenue
Highveld Technopark
Johannesburg
Gauteng, South Africa

AI Asia Pacific Institute offices

Asia

Singapore

10 Anson Road #32-02,
International Plaza
079903

North America

San Francisco

28 Geary Street, Suite 650
San Francisco, CA 94108