

RETHINKING DIGITAL SAFETY

Why online harms should not be inevitable in a digitalising world.

Once thought to be individual responsibility, online harms have been reframed to be a product of how platforms are designed, governed and socially sustained.

The speed of response to online harms is critical; the value of relief diminishes if harms continue to circulate while victims seek redress.

Research plays a key role in enabling policy response to adapt to new forms of harm as technology evolves.

As societies worldwide digitalise, it is almost commonplace for online harms to be treated as the equivalent of background noise—unpleasant, sometimes distressing, but ultimately part and parcel of the digital realm. Trolls emerge, harassment occurs and platforms moderate, though with minimal effort.

Emboldened by anonymity and enabled by sharing platforms, trolls—users who deliberately antagonise others on the Internet—can engineer abuse at scale. Women have been especially vulnerable, with substantial exposure to grossly sexist and misogynistic content online. Social media platforms like Facebook or Telegram, or discussion forums such as Quora or Reddit, or video sharing apps including YouTube or TikTok host ‘lawful but awful’ content that demean and objectify women through the vicious tropes of toxic masculinity.

Other vulnerable groups such as ethnic and religious minorities, LGBTQ+ individuals, migrants and persons with disabilities often face disproportionate levels of online harm. These include hate speech, coordinated harassment, doxxing, which involves publishing personal information to inflict harm, misinformation and algorithmic discrimination, all of which can exacerbate marginalisation offline.

Although the Internet has undoubtedly democratised online participation, it nevertheless reproduces and even amplifies socio-economic inequalities around class, gender and ethnicity. For too long therefore, the prevailing assumption, implicit if not explicit, has been that individuals should just grit their teeth when facing online annoyances, or simply log off.

That assumption no longer holds in Singapore.

In this highly connected city-state, as in a growing swathe of digitally advanced societies, online harms have undergone a notable and decisive reframing. They are no longer regarded as episodic incidents, but as unwelcome byproducts of the online environment that undermine trust, impede participation and reverse inclusion. This shift in understanding has driven an equally significant evolution in policy, moving from reactive, case-by-case interventions to a more systemic and anticipatory approach.

Far from a niche regulatory issue, online harms and their prevalence reflect how well digital ecosystems function, who feels empowered within them and, ultimately, whether trust is both intuited and experienced.

THE TURN TO COLLECTIVE RESPONSIBILITY

A defining feature of Singapore’s approach has been its refusal to treat online harms as solely a legal problem. The formation of the Sunlight Alliance for Action (AfA) marked an early recognition that the issue is fundamentally social before it is regulatory.

Launched by the Ministry of Communications and Information (precursor to the Ministry of Digital Development and Information) in July 2021, the AfA sought to address gaps in digital safety and strengthen support systems for victims of online harms. Co-chaired by then Senior Minister of State Ms Sim Ann and Senior Parliamentary Secretary Ms Rahayu Mahzam, the AfA brought together 48 representatives from the people, public and private sectors. As a whole-of-society initiative, it also contributed to the broader national Digital for Life movement that promotes digital inclusion.

With this diversity of representation, the AfA and its string of outreach efforts helped surface the “hidden majority” of experiences that rarely make it into formal complaints or legal cases. It made visible what

had previously been normalised or dismissed and, just as importantly, redistributed responsibility. Online harms were no longer framed as a failure of individual resilience, but as the outcome of system design, platform governance and social norms.

With this critical reframing, the locus of intervention shifts. It is no longer sufficient to punish bad actors after the fact. The question becomes: how can digital environments be redesigned such that harm is less likely to occur, less likely to spread and easier to address when it does? The objective is not merely to respond to incidents, but to reduce systemic vulnerability.

WHEN HARM BECOMES AMBIENT

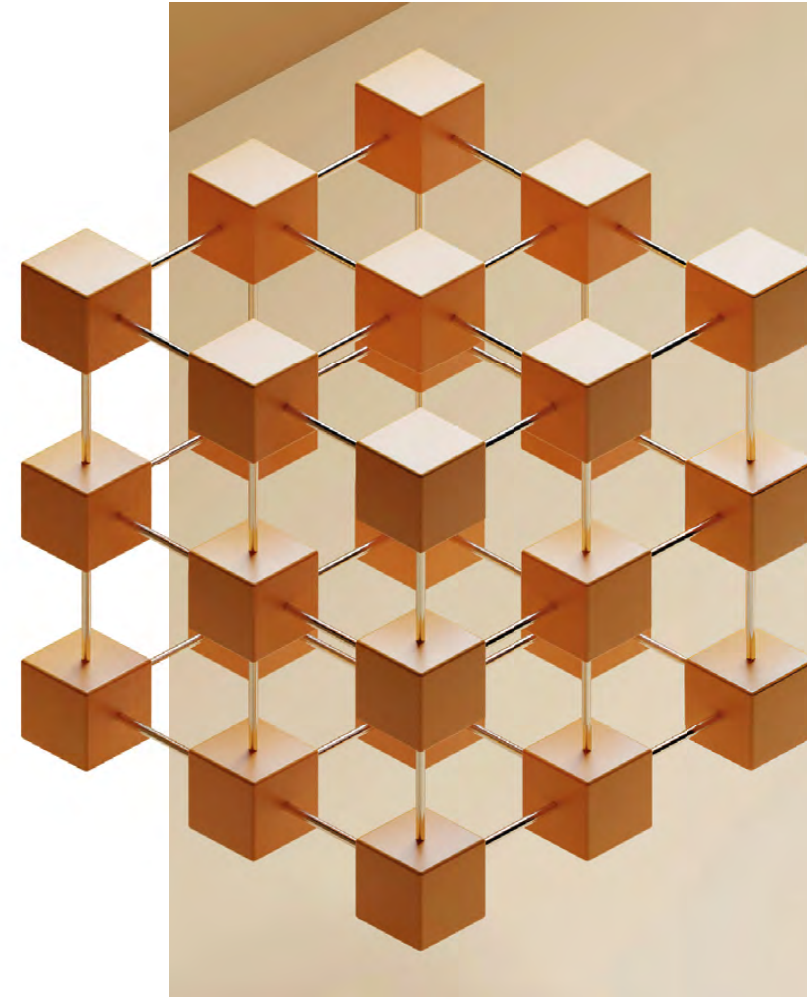
Indeed, one of the more insidious aspects of online harm is not how extreme it is, but rather its pedestrian nature. Much of it does not rise to the level of criminality, instead residing in a grey zone of persistent harassment, non-consensual sharing and coordinated pile-ons that cumulatively erodes a person’s sense of safety.

In Singapore, this has become increasingly visible in both research and lived experience. Victims do not necessarily seek punishment; they seek cessation. They want the content taken down, the abuse to stop and the ability to reclaim their digital presence. Yet for many, the pathways to redress have historically been difficult to navigate. As policymakers themselves have acknowledged, victims are often deterred by existing processes and struggle to identify perpetrators or secure timely relief.

Clearly, harm resides not only in the content, but in the system’s inability to respond at the speed and scale at which harm propagates. Singapore’s policy evolution thus reflects adaptation to this sobering reality.

KEY LEGISLATIVE SHIFTS PERTAINING TO ONLINE HARMS

Singapore’s legislative trajectory reflects this shift from episodic enforcement to systemic design. The Protection from Harassment Act (POHA) that was passed in 2014 was an important starting point. It established that harms such as harassment, stalking and doxxing are not trivial but actionable. Yet it was, in many ways, a product of an earlier Internet where harms were more discrete and perpetrators more identifiable.



Online harms are no longer framed as a failure of individual resilience, but as the outcome of system design, platform governance and social norms.

As digital environments became richer and more complex, the limitations of this approach became clearer: harm could spread faster than legal processes could contain it; perpetrators could remain anonymous or operate across jurisdictions; and platforms, which mediated much of this activity, sat somewhat ambiguously between facilitator and gatekeeper.

The Online Criminal Harms Act (OCHA) passed in 2023 then marked a decisive departure from this earlier model. It introduced the ability for authorities to intervene upstream, issuing directions to platforms and other entities to disrupt harmful activity before it fully materialises. With this conceptual shift, the law does not merely adjudicate wrongdoing after it occurs but rather seeks to influence conditions under which wrongdoing can occur at all.

This logic is extended further in subsequent developments. On 18 July 2023, the Infocomm and Media Development Authority introduced the Code of Practice for Online Safety, requiring platforms to establish systems and processes to mitigate harm, inscribing accountability into the architecture of digital services. The emphasis is increasingly on outcomes rather than prescriptions, allowing for flexibility while maintaining clear expectations.

RECENTRING THE VICTIM

If those preceding legislations reflected a move toward systemic intervention, the Online Safety (Relief and Accountability) Act (OSRA) passed in November 2025 represents a shift toward the lived experiences of victims.

One of the most persistent critiques of earlier frameworks was that they were, in practice, inaccessible. Although legal remedies existed, they were often too slow, too complex or too costly for those affected. OSRA addresses this gap directly by reorienting the system around the needs of victims. Importantly, it recognises a simple but powerful reality: in the context of online harm, speed is justice. The value of a remedy diminishes rapidly if harm continues to circulate unchecked.

The Act therefore introduces mechanisms designed to deliver timely and practical relief. It establishes statutory torts that clarify responsibilities across users, administrators and platforms, and enables victims to seek redress more directly. It also addresses the problem of anonymity by allowing for the identification of perpetrators under appropriate safeguards.

Perhaps most significantly, OSRA creates the institutional foundation for a new mode of governance through the Online Safety Commission.

ESTABLISHMENT OF THE ONLINE SAFETY COMMISSION

The establishment of the Online Safety Commission (OSC) in June 2026 signals a maturation of Singapore’s legislative approach. Rather than relying solely on courts or dispersed regulatory powers, the Commission was conceived as a dedicated node within the governance architecture. Taking learnings from Australia’s eSafety Commissioner, the OSC is intended to operate as a first point of recourse, capable of acting quickly and proportionately. It can issue directions to remove content, restrict accounts or require other remedial actions, targeting not only perpetrators but also platforms and administrators.

In doing so, it addresses a fundamental mismatch that has long characterised online harm: the speed of harm versus the speed of response. Traditional legal systems are not designed for real-time intervention, but a specialised body such as the OSC is. This is consistent with broader global trends, where jurisdictions are experimenting with dedicated regulators for digital harms, but Singapore’s approach is notable for the degree to which it integrates this institution within a broader ecosystem of laws, codes and partnerships.

This has tangible business consequences as online harms go directly to the question of trust. Digital platforms, marketplaces and communities depend on users’ willingness to participate, share and transact. When environments are perceived as unsafe, participation becomes selective, guarded or withdrawn altogether.

It affects user growth, engagement quality, brand reputation and, ultimately, revenue. More subtly, it shapes who participates: if certain groups—women, minorities or younger users—are disproportionately affected by harm, their absence or silence distorts the very markets and communities that businesses seek to serve.

With online harms regulation, businesses are impelled to invest in governance capabilities, from content moderation systems to internal escalation processes, and engage more continuously with regulators. Ultimately, a healthier information landscape will be win-win for consumers and businesses alike.

SETTING THE RESEARCH AGENDA

Undergirding Singapore’s approach is a less visible but equally important thread: academic research.

Policy in this domain cannot rely on intuition or online virality alone. Online harms are dynamic, context-dependent and often difficult to measure. Research provides the empirical grounding needed to understand not just the prevalence of harm, but its forms, trajectories and impacts.

In the case of OSRA, for instance, the emphasis on victim-centric remedies reflects extensive consultation and evidence-gathering. Policymakers did not simply assume what victims needed but engaged the ground through expert and public engagement dialogues.

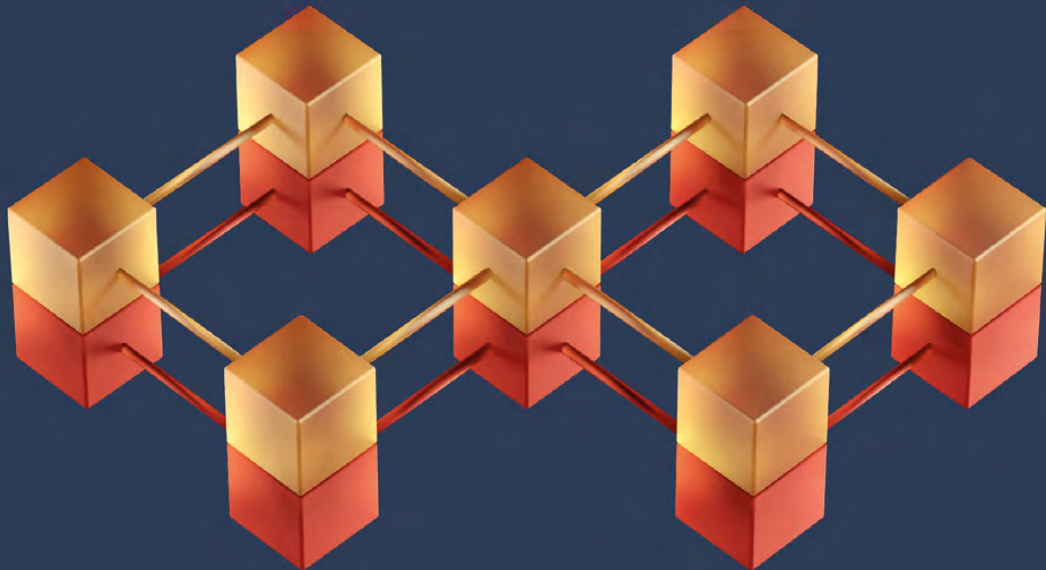
Research also plays a critical role in anticipating what comes next. The rise of generative AI, for example, is already reshaping the landscape of online harm, enabling more sophisticated forms of impersonation, manipulation and synthetic content. Without sustained research, policy risks always being one step behind. To this end, the Research Workstream of the Sunlight AfA, which I had the honour of leading, mapped out a research agenda that guides future work on this fast-evolving area, identifying key issues and actors, impact assessment approaches as well as appropriate research methods.

Issues	
Conceptual • <u>Definitions</u> of harm and severity of harms: provide clarity and impact on reporting and responses by companies, criminal liability, reporting and enforcement, proportionality of response • <u>Nomenclature</u>: naming different kinds of online harms in more value-neutral and accurate manner e.g. alternatives to ‘revenge porn’	
Individual • <u>Victims</u>: experiences, support needed, wellbeing (physical and mental) • <u>Bystanders</u>: diffusion of responsibility, bystander intervention • <u>Accomplices</u> who aid and abet online harms through gatekeeping role(s) • <u>Resilience of individuals</u> in face of online harms • <u>At-risk individuals</u> less inhibited about perpetrating online harms and interventions required • <u>Family structures and dynamics</u>’ impact on individual resilience, combined with issues of class, household composition etc.	Group/Community • <u>Group norms</u> in different online platforms, e.g. Singaporeans more reticent about calling out bad behaviour? • <u>Social resilience</u> as manifested in support from social groups and networks for victims in terms of direct assistance or other resources; issues of resource capacity and nature and scope of support • <u>Religious/cultural groups</u>: social stigma, community and religious leadership, religious discourse promoting or preventing online harms (comparative research across religions). Cultural groups could encompass ethnic, youth, pop cultures etc. • <u>Family support</u> or lack thereof has implications for community response • <u>Groups of interest</u>: <i>domestic workers</i> (vulnerable), <i>families</i> (need guidance on parental mediation), <i>young males</i> (exposure to porn, normalisation of attitudes towards sexual relationships)
Industry (technology, media) • <u>Best practices</u>: industry engagement for greater knowledge sharing and industry self-regulation including industry codes of practice • <u>Safety features</u> introduced and their efficacy (media and tech platforms) • <u>Reporting systems</u> of different tech platforms and their efficacy (comparative analysis) • <u>CSR initiatives</u> and their value • <u>Guidelines on privacy protection/codes of professional practice</u> in reporting sexual incidents • <u>App design and metrics</u> that promote hostile online environments e.g. dehumanisation, monetisation, use of avatars	Policy/Advocacy • <u>Codes of practice</u> for online safety: comparative analysis of diverse global approaches • <u>Dedicated enforcement agency</u> for online harms: feasibility analysis • <u>Evidence-based research</u> to drive recommendations • <u>Mainstream education components</u> to raise students’ understanding of online harms and their responsibility and agency (up to and including tertiary education)
Impact Assessment • <u>Periodic assessments</u> of the effectiveness of public education, legislation, technology company responses • <u>Perceptions</u> of different respondent profiles and trends over time • <u>Longitudinal analysis</u> of impact of policies	

Approaches
Qualitative, ethnographic research • Interviews, focus groups, discourse and conversational analysis, case-based research, online ethnography (rich contextualised findings on victim trauma, bystander attitudes and interventions etc.)
Quantitative, computational research • Panel study to follow a pool of participants for assessing longitudinal effects • Developing accurate survey measures for sensitive questions relating to online harms (tracking non-response rate to particular questions, by different respondent profiles etc.) • Machine learning supported content analysis of videos and discussion forum threads with harmful content (trends in engagement with misogynistic content, common tropes etc.)
Technological tool development • Tools to help detect misinformation, collect evidence in claim verification process, analysing social media content to see if images have been repurposed and taken out of context, tools to construct profiles of people to help verify identity of persons to potentially befriend, tools to guide prosocial behaviour

Summary of key issues and actors, impact assessment and approaches to research developed by the Research Workstream of the Sunlight Alliance for Action (AfA).

Resilience at the level of a system is not a problem to be solved once and for all, but a condition to be constantly managed.



Equally important is the role of research in translation. Academic insights must be made legible to business, policymakers and the public. It is not merely a communication challenge but a governance imperative: a society that better understands online harms is better equipped to respond to them. In this regard, Singapore Management University’s Yong Pung How School of Law and its Centre for Digital Law drove a landmark event that demonstrated the importance of tackling online harms. Organised with Ministry of Law Singapore, the Online Harms Symposium: Exploring New Solutions and Legal Remedies for a Safer Online World ran from 25-27 September 2023.

The symposium convened leading local and international experts on online safety, including legal professionals, academics, community organisations and technology companies, to discuss emerging challenges and explore effective and accessible solutions to online harms. Attracting over 160 participants each day, the event fostered lively and substantive discussions among stakeholders from the public, private, and people sectors, including major technology firms such as Amazon, Apple, ByteDance, Google, LinkedIn, Meta and X. Its accompanying report serves as a pithy guide to the critical issues pertaining to online harms.

GOVERNING A MOVING TARGET

If there is a single lesson from Singapore’s experience, it is that online harms cannot be addressed through static solutions. They evolve alongside technology, often in counterintuitive ways. The policy response too, therefore, must be adaptive. It must combine legal frameworks with institutional capacity, technological tools with human judgement and regulation with education. It must also remain attentive to trade-offs, particularly the balance between safety and freedom of expression. What has changed in Singapore is not simply the set of laws, but the underlying mindset. In the case of online harms, they must be understood as systemic risks that require coordinated, sustained and evidence-based response involving a broad range of stakeholders. Universities are poised to play a vital role in engaging and even convening these stakeholders, while undertaking nuanced research that can shape policy making and public outreach. This, ultimately, is what resilience looks like at the level of a system: not a problem to be solved once and for all, but a condition to be constantly managed.



SUN SUN LIM is Vice President, Partnerships & Engagement and Lee Kong Chian Professor of Communication and Technology at Singapore Management University. She has researched extensively the social impact of technology with over 100 publications including *Transcendent Parenting - Raising Children in the Digital Age* (2020) and articles in flagship journals like *Nature* and *Journal of Computer Mediated Communication*. She was a Nominated Member of the 13th Parliament of Singapore and a Top 50 Asia Women Tech Leaders 2024 honouree. She has a PhD from the London School of Economics and an Honorary Doctorate from the University of Copenhagen.