

Justice beyond virality

How platform governance, design features and incentive structures affect redress in online spaces



CERT

POLICY PAPER

Centre for Responsible Technology

The Centre for Responsible Technology (CERT) is an autonomous research centre established by the Institute of Strategic & International Studies (ISIS) Malaysia with the foundational support of the Malaysian Communications and Multimedia Commission. CERT's mission is to provide the government and key stakeholders with the knowledge and insights needed to shape technological development guided by the principle of responsibility, ensuring that such development prioritises public interest, human rights, accountability, and effective regulation.

CERT's research mandate is intersectional, focusing on the relationship between technology and four primary domains: society, sustainability, safety, and security. Through research, dialogue and stakeholder engagement, CERT contributes to informed policy discussions on emerging technologies and their societal implications, working at the intersection of academia, industry and government.

[POLICY PAPER]

Justice beyond virality

How platform governance, design features and incentive structures affect redress in online spaces

By

Dr Gayathry Venkiteswaran

Foreword by

Datuk Prof Dr Mohd Faiz Abdullah

Illustrated by

Ashikin Hussin

[AUTHOR]

Dr Gayathry Venkiteswaran is an assistant professor at the University of Nottingham Malaysia, where she teaches and researches on media and politics from the perspectives of gender justice, political economy, democracy and human rights. She has over 25 years of experience in journalism, and advocacy for media freedom and freedom of expression in Malaysia and Southeast Asia.

[ACKNOWLEDGEMENTS]

The author would like to thank Claudian Navin Stanislaus, Nalini Elumalai, Quin Leong, Farhan Shahmi Abdullah, Dr Abszra Davadason, Dr Benjamin YH Loh, Alyaa Alhadjri and Syed Jaymal Zahiid, for their insights and advice in the preparation of this policy brief, and Tisha Daniea Muhammad Adam and Elynn Ong for their assistance.

Foreword

"[it is] of fundamental importance that justice should not only be done, but should manifestly and undoubtedly be seen to be done," reads the prominent legal maxim introduced by Lord Hewart CJ in *R v Sussex Justices ex parte McCarthy*.

The judgment highlights how justice cannot merely be reduced to procedures and technicalities, but must also satisfy broader expectations of fairness, legitimacy and public confidence. This maxim, however, can be turned on its head when justice itself is only seen to be done rather than having gone through the proper procedures.

Put another way, the pursuit of justice can give way to a form of mob justice, where the views of the majority, or those with the loudest voices, take precedence over the black letter of the law. Followed to its natural conclusion, visibility then becomes conflated with accountability, while public condemnation is mistaken for due process.

This paper situates this risk within the broader communications landscape, where social media use is ubiquitous and where users are increasingly turning to digital platforms to seek redress, expose wrongdoing and hold individuals or institutions to account. To be clear, this growing reliance on social media as a mechanism of accountability did not emerge in a vacuum. It reflects, in part, public frustrations with existing institutions and a desire for more immediate forms of recourse.

Yet these same platforms were not designed to function as courts, regulators or public ombudsmen. Their primary purpose is to facilitate engagement, and the incentives underpinning them do not necessarily align with the principles of fairness, transparency or procedural justice.

As a result, important questions arise. What happens when accountability becomes dependent on virality? Who determines which grievances deserve attention and which are ignored? And what are the implications when governance functions are increasingly mediated through private companies that are neither publicly elected nor democratically accountable?

This paper contributes to a growing discussion on the relationship between technology, power and public institutions. In doing so, it invites us to reconsider whether the pursuit of digital justice should be measured by how visible it is, or by whether it ultimately strengthens the foundations of justice itself.

Datuk Prof Dr Mohd Faiz Abdullah

Executive Chairman

Institute of Strategic & International Studies (ISIS) Malaysia

Abbreviations

AI	Artificial intelligence
CMA	Communications and Multimedia Act
DSA	Digital Services Act (European Union)
MCMC	Malaysian Communications and Multimedia Commission
SiSPAA	Public Service Complaint Management System (Sistem Pengurusan Aduan Awam)
UDHR	Universal Declaration of Human Rights
UN	United Nations

Executive summary

- Rapidly expanding social media platforms increasingly function as means of pursuing digital justice through online visibility, deployed as parallel enforcement mechanisms at the expense of formal bureaucratic systems.
- Such visibility is mistaken for justice or symbolic recognition, despite often failing to address underlying structural and systemic problems. Further, unlike public mechanisms, platform-driven systems are governed by private corporations, opaque algorithms and engagement-based business models that are not democratically accountable.
- Platformised governance reshapes accountability ecosystems through three interrelated dynamics. First, social media platforms replace public institutions as informal grievance and redress systems. Second, platform architectures incentivise virality, outrage, emotional amplification and public shaming as means of achieving justice. Third, the increasing reliance on platform reporting systems and algorithmic moderation transfers governance towards unaccountable private companies.
- Therefore, Malaysia requires a rights-based, transparent and accountable digital governance framework that strengthens public accountability while addressing the structural dynamics of platform power.
- However, policy responses currently remain largely content- and enforcement-oriented, rather than critically examining systemic risks embedded within platform design and recommendation systems. This risks collateral censorship, politicised enforcement and weakened democratic oversight.
- Recommendations include independent algorithmic audits, mandatory transparency reporting, stronger procedural safeguards, independent oversight bodies, interoperable grievance systems and greater public investment in institutional accountability mechanisms.

1. Introduction

Digital platforms have become central infrastructures for communication, participation and public accountability in Malaysia. Citizens increasingly rely on social media platforms to raise grievances, document institutional failures, pressure authorities and mobilise public support. Complaints regarding road infrastructure, consumer disputes, workplace abuse, gender-based violence, bullying, religious tensions and political controversies, for instance, are routinely channelled through social media rather than formal institutional pathways. The increasing visibility of these practices reflects broader transformations in governance, public trust and the relationship between citizens and institutions, marking a broader shift away from formal democratic accountability towards profit-oriented platform logics.

This shift is not occurring in isolation. Globally, scholars have warned that digital platforms increasingly function as systems, or infrastructures, of governance, rather than merely communication intermediaries. Gillespie argues that platforms actively shape public discourse through moderation policies, recommendation systems and visibility controls.¹ Suzor similarly notes that digital platforms exercise constitutional power by establishing rules, adjudicating disputes and enforcing behavioural norms across transnational publics.² These developments have contributed to what some scholars describe as *platformised governance*, where private companies increasingly regulate speech, participation and access to remedy.³

In Southeast Asia, these transformations intersect with longstanding governance challenges, uneven institutional trust and rapidly expanding digital participation. Malaysia presents an important case study because of its high social media penetration rate, strong mobile connectivity and politically contested information environment. Citizens frequently experience fragmented, bureaucratic, inaccessible or ineffective formal accountability systems. Complaint mechanisms may exist across ministries and agencies, but navigating these often require institutional literacy, time and persistence that many citizens do not possess. In contrast, social media platforms provide immediacy, visibility, emotional validation and the possibility of rapid response.

The attraction of social media as a grievance mechanism therefore reflects deeper structural conditions. Individuals often perceive that institutions only respond when issues become public, embarrassing or politically costly. Viral exposure functions as an informal escalation mechanism capable of compelling authorities to action. This creates a feedback loop in which *public visibility becomes central to accountability*. However, such visibility is not synonymous with justice. Social media platforms operate according to commercial logics structured around optimising engagement, extracting data and maximising attention. Algorithmic systems reward emotionally charged, polarising and sensational content because such material generates interaction and prolongs user engagement. Consequently, outrage and virality are structurally incentivised within platform ecosystems. Grievances framed in emotionally charged terms are more likely to circulate

widely compared to nuanced or procedural discussions. Rather than achieving justice or symbolic recognition, visibility on platforms often fails to resolve underlying structural and systemic problems.

These dynamics reshape not only communication methods, but also *political agency*. Algorithmic curation determines what information becomes visible, whose grievances are amplified and which narratives gain legitimacy. Platforms increasingly function as invisible editors of public discourse. Yet unlike traditional public or mass media institutions, these systems often lack transparency, democratic oversight or procedural safeguards.

The growing dependence on platform reporting systems also contributes to a broader displacement of official accountability. Citizens are increasingly directed towards platform complaint mechanisms even in cases involving harassment, abuse or harmful conduct. Responsibility is shifted away from public institutions and towards private moderation systems that remain opaque and inconsistently enforced. Platform reporting systems may provide immediate channels for flagging harmful content, but rarely fairness, due process, transparency or sustainable institutional reform.

This brief's central question is: *how is (digital) justice governed in a platform society?* Specifically, it explores how platformised governance reshapes Malaysian accountability ecosystems and its implications for democratic participation, public trust and institutional legitimacy. Informed by desk research, comparative regulatory developments, civil society reports, policy analysis (including platform policy documents), selected local case studies and a roundtable with experts, it draws upon Malaysian and international perspectives to illustrate broader patterns of platform-mediated grievance and accountability.

Section 2 reviews the literature on digital justice, platformised governance and accountability; Section 3 examines international standards around platformised governance and the local landscape; Section 4 presents case examples and issues surrounding digital justice; and Section 5 presents recommendations for stakeholders, including, but not limited to, independent algorithmic audits, mandatory transparency reporting, stronger procedural safeguards, independent oversight bodies, interoperable grievance systems and greater public investment in institutional accountability mechanisms.

2. Literature review

Digital platforms enable citizens to mobilise collectively, document abuses and pressure authorities, through their participation in various forms of *decentralised accountability*. Indeed, a growing body of literature conceptualises digital vigilantism as one of its key manifestations. According to Legocki et al., it is a form of consumer voice, where publics use social media to pressure institutions following perceived failures.⁴ The concept of *digital justice* thus emerges at the intersection of platformised governance, user participation and evolving accountability practices. But digital justice is inherently

ambivalent, as illustrated by the rise of online naming-and-shaming cultures. The #MeToo movement for example, demonstrated how social media could expose systemic gendered and sexual abuses ignored by institutional structures. Survivors used digital platforms to document their experiences, warn others and pressure institutions into responding.

Yet such practices may blur distinctions between accountability and punishment, particularly when allegations circulate without procedural safeguards. Venkiteswaran's work on hatred and freedom of expression in Asia highlights how social media simultaneously enables mobilisation against injustice while facilitating mob attacks, harassment and coordinated intimidation campaigns.⁵ Loveluck identifies multiple forms of digital vigilantism (including flagging, hounding, investigating and organised leaking) which rely on collective participation, networked outrage and public denunciation.⁶ Similarly, Cox demonstrates how platform affordances (such as sharing and various forms of reacting) and algorithmic amplification enable the quick mobilisation of affective publics around claims to justice, as seen in cases like the #JusticeForChristineDacera campaign in the Philippines.⁷ Users engaged in widespread naming-and-shaming and speculation, but the episode devolved into misinformation and reputational harm for individuals wrongly implicated, illustrating the dangers of weaponised data and premature judgment. As Cox writes:

While this type of censoring and surveillance is believed to positively contribute to a transparent society as a form of engagement and empowerment, there is such potential for an unregulated community punishment which encourages trial by social media.⁸

The repercussions are especially evident in high-profile global cases. The 2013 Boston Marathon bombing investigation in the United States of America provides a good example of crowdsourced justice gone wrong. Reddit users and social media participants collectively attempted to identify suspects, leading to the misidentification and harassment of innocent individuals, including Sunil Tripathi, whose family received threats after his false accusation of involvement. Unfounded allegations and damaged reputations ultimately undermine both justice and public trust, illustrating how digital vigilantism can rapidly escalate into online witch hunts, particularly in the absence of institutional safeguards.⁹ Further, it often emerges when institutional trust is weak, or where publics perceive formal accountability mechanisms as being inaccessible or ineffective.

In short, although digital vigilantism may seemingly strengthen public participation and engagement, it actually destabilises procedural fairness and institutional legitimacy – in other words, it undermines fundamental pillars of the democratic system. These examples demonstrate how seemingly effective digital justice systems frequently prioritise immediacy and visibility over verification or due process. Social media platforms are structurally designed to reward engagement, meaning that content which provokes outrage, anger, fear or moral condemnation tends to circulate more widely because algorithmic systems optimise interaction and retention. Outrage amplification is not incidental, but deeply embedded within platform business models.

Consequently, platformised governance cannot be understood solely through the lens of individual expression. The broader architecture of visibility facilitated by recommendation systems, trending mechanisms, engagement metrics and moderation practices shape public discourse in ways that remain largely opaque to users. Further, those with greater digital literacy, stronger networks, emotional storytelling capacity, or access to influential amplification channels possess greater ability to mobilise attention. In short, digital justice operates through deeply unequal infrastructures of visibility.

These dynamics are closely linked to broader debates surrounding digital constitutionalism. Celeste et al. argue that platforms increasingly operate as quasi-legal systems exercising constitution-like authority over digital publics.¹⁰ After all, platforms establish rules governing acceptable behaviour, determine enforcement mechanisms, adjudicate disputes and implement sanctions affecting billions of users globally. Unlike democratic institutions, platforms are not grounded in constitutional accountability, public participation or procedural transparency. Decisions are often shaped by commercial incentives, public relations considerations, geopolitical pressures and risk-management priorities, rather than in relation to rights-based standards.

This creates significant governance dilemmas. On one hand, platforms face pressure to remove harmful content quickly and aggressively. On the other, overly broad moderation risks suppressing legitimate expression, political dissent and minority voices.¹¹ The challenge is particularly acute in politically sensitive environments, where state authorities may pressure platforms to remove critical content under the pretext of maintaining public order or national harmony.

Ultimately, digital justice is neither inherently emancipatory nor inherently harmful. It is conceptually paradoxical: potentially positive democratic implications depend heavily on undemocratic platform architectures, opaque governance frameworks, institutional capacities and broader political conditions. Addressing these tensions requires moving beyond simplistic debates about content moderation, and towards *systemic accountability approaches* capable of addressing structural power asymmetries within platform ecosystems. It also raises questions about what it means for users to engage in *digital citizenship*, a concept developed more thoroughly globally, which is broadly defined as the capacity to actively, responsibly and continuously participate in social and civic activities, online.¹²

3. Overview of platformised governance approaches

3.1 International digital standards

International debates on digital governance increasingly recognise that platform accountability cannot be addressed solely through reactive content moderation. Instead, regulatory attention has shifted towards addressing the systemic risks embedded within platform architectures, recommendation systems and commercial business models. This is particularly crucial when platform moderation decisions increasingly affect political participation (by exercising substantial influence over public discourse and information flows), public visibility, professional livelihoods and social reputation.

Rights-based approaches remain central to international digital governance frameworks. Freedom of expression, which is protected under the United Nations (UN) – Article 19 of the Universal Declaration of Human Rights (UDHR) and the International Covenant on Civil and Political Rights – applies fully in digital environments. International standards emphasise that restrictions on speech must remain lawful, necessary, proportionate and subject to procedural safeguards. However, digital environments raise significant concerns regarding due process and access to remedy. Users often receive limited explanations regarding enforcement decisions and encounter weak appeals mechanisms. The UN special rapporteur on freedom of expression has repeatedly warned that opaque algorithmic systems and private moderation practices can undermine democratic accountability and restrict freedom of expression without adequate oversight.¹³

Consequently, the establishment and maintenance of platformised governance standards are increasingly critical. The Santa Clara Principles on Transparency and Accountability in Content Moderation (first formulated in 2018) are a key reference, emphasising procedural safeguards such as the need for notice, explanations, appeals processes and transparency reporting.¹⁴ Similarly, the UN Guiding Principles on Business and Human Rights, especially Principle 31, call upon companies to observe due diligence with regard to human rights, including assessing risks related to content moderation, surveillance and algorithmic bias. They must also establish operational-level grievance mechanisms that are legitimate, accessible, predictable, equitable, transparent and rights-compatible.

3.2 Comparative approaches to regulation

Globally, regulatory models are evolving. One important trend across the following frameworks is the shift of regulatory attention away from mere content moderation towards a focus on systems. This suggests that policymakers increasingly recognise how harmful outcomes are often systemically embedded within platform architectures optimised for engagement and attention-extraction. Notable examples are as follows:

- The European Union Digital Services Act (DSA) represents one of the most comprehensive attempts to regulate systemic platform risks.¹⁵ Rather than focusing exclusively on illegal content, the DSA requires that very large online platforms assess and mitigate systemic risks relating to disinformation, electoral integrity, mental health harms, gender-based violence and impacts on fundamental rights.
- Importantly, the DSA also introduces obligations relating to transparency, independent auditing, researcher access and algorithmic accountability. This reflects a broader recognition that platform harms often arise not only from specific content, but also underlying amplification systems and commercial incentives.
- Australia's Online Safety Act adopts a somewhat different approach by focusing primarily on harm mitigation.¹⁶ This legislation empowers the eSafety Commissioner to issue removal notices and address cyberbullying, image-based abuse and harmful online conduct. While this framework emphasises user protection, critics warn that broad enforcement powers may create risks for freedom of expression.
- The United Kingdom's Online Safety Act similarly adopts a duty-of-care model requiring platforms to proactively manage risks.¹⁷ However, in parallel with concerns in Australia and elsewhere, civil society organisations have raised concerns that broad definitions of harmful content may incentivise over-removal and collateral censorship.
- South Korea's approach highlights additional dimensions of platformised governance by focusing on market fairness, algorithmic transparency and protections for smaller business users dependent on digital platforms. It combines external and self-regulation, while recognising that platform governance extends beyond the regulation of expression and into broader economic considerations.¹⁸

Across these jurisdictions, digital governance is understood as neither relying solely on voluntary corporate self-regulation nor state-centric models. Indeed, allocating regulatory responsibility to states alone can create additional risks. Governments may pressure platforms to remove politically sensitive content or comply with broad surveillance demands, for example. In contexts where legal safeguards remain weak, platforms may then engage in preventive over-moderation to minimise

regulatory risk. This creates conditions for collateral censorship, whereby legitimate political speech or dissenting viewpoints are suppressed indirectly through internal platform risk-management strategies. Public accountability mechanisms therefore risk being subordinated to opaque negotiations between states and corporations.

3.3 Platform (self-)governance and digital constitutionalism

Major companies, including Meta, X (formerly Twitter), TikTok and YouTube, have established rules governing acceptable speech, determined enforcement procedures, moderated public discourse and shaped visibility through algorithmic systems. Platforms publicly frame these moderation systems in the language of safety, community standards and human rights.¹⁹ Reporting systems allow users to flag harmful content related to harassment, hate speech, misinformation or abuse. Transparency reports provide aggregated data regarding enforcement actions and government requests. Appeals systems and oversight mechanisms are presented as safeguards to promote accountability. Amidst these reforms, a key theme is increasingly emerging: platforms are expected to act as *providers of remedy* and not merely moderators of content. Mechanisms such as appeals processes, transparency reports and independent oversight bodies (for example, Meta's Oversight Board) are attempts to operationalise this principle, although concerns about independence and effectiveness remain salient.²⁰ Substantial gaps remain between public commitments and actual enforcement. Civil society organisations including ARTICLE 19, Ranking Digital Rights and PEN America have consistently identified significant weaknesses pertaining to transparency, consistency, procedural fairness and access to remedy.²¹

The automation of moderation systems complicates these challenges.²² Platforms now rely on machine learning systems and automated detection tools to manage enormous volumes of content. While potentially improving scalability, significant risks include contextual misinterpretation, discriminatory enforcement and over-removal of content. Marginalised communities often experience disproportionate harms under these systems, which inadequately protect vulnerable groups while simultaneously penalising activists, journalists or minority voices documenting harmful conduct. Powerful actors, meanwhile, may receive preferential treatment because of political or commercial considerations. Studies on the policies drafted by Meta, YouTube and X have shown that these texts often leave marginalised groups under constant threat of removal, while platform practices typically protect powerful groups and influential accounts that tend to be the main drivers of online and offline harms.

Essentially, the use of the language of human rights in relation to social media platform policies on content moderation practices is largely nominal, while full integration of their ideals into platformised governance is deferred.²³ These problems cannot be separated from platform business models, which operate within surveillance capitalist frameworks dependent on data extraction, behavioural

prediction and engagement optimisation. Outrage, controversy, emotional intensity and polarisation are not merely unfortunate side effects, but structurally profitable conditions. Algorithmic systems therefore prioritise content that is likely to maximise interaction, where recommendation systems amplify emotionally charged material to sustain user attention.

In these environments, outrage-driven mobilisation receives disproportionate visibility. Platforms may tolerate inflammatory or polarising content where it drives engagement, while aggressively moderating material that threatens advertiser interests or regulatory relationships. The increasing dependence on user reporting systems transfers substantial governance responsibilities onto publics themselves. Users become unpaid moderators responsible for flagging harmful conduct, documenting abuse and navigating opaque complaint processes. Yet users still receive limited explanations regarding why content is removed, accounts restricted or complaints rejected under these moderation systems. Appeals processes may exist formally, but often remain difficult to navigate or are ineffective in practice. PEN America's research demonstrates that users frequently encounter deeply flawed abuse reporting systems characterised by limited feedback, inconsistent enforcement and inadequate support.²⁴ Marginalised communities facing coordinated harassment often struggle to obtain timely responses despite repeated reporting efforts.

These governance problems illustrate broader tensions surrounding digital constitutionalism. Terms of service effectively operate as governing constitutions shaping participation rights, enforcement mechanisms and access to visibility. Platforms effectively function akin to private legal systems governing public discourse, but without equivalent democratic accountability obligations. Procedural safeguards remain weak compared to public legal systems. There is limited transparency regarding decision-making criteria, inconsistent opportunities for appeal and little meaningful external oversight: independent oversight boards or advisory councils remain limited in scope and authority. In cognisance of the limitations of both state responses and platform self-governance, effective governance arguably requires *hybrid approaches* combining human rights protections, independent oversight, transparency obligations, procedural safeguards and systemic accountability mechanisms.

3.4 Malaysia's digital landscape and governance

Malaysia's digital landscape is characterised by near-universal connectivity, widespread mobile device ownership and intensive social media engagement, but also expanding state regulation and concerns regarding digital rights, censorship and online expression. According to DataReportal, Malaysia recorded 34.9 million internet users in early 2025, representing an internet penetration rate of 97.7% of the population, alongside 43.3 million active mobile cellular connections (equivalent to 121% of the population, driven by multiple device usage and SIM card ownership).²⁵ There were an estimated 25.1 million social media user identities, or approximately 70% of the population, indicating the centrality of digital platforms to everyday political communication, commerce and civic engagement. Key platforms (such as YouTube, TikTok, Instagram, Facebook and X) are dominant spaces for the articulation of public discourse, news circulation and political participation.

The Malaysian Communications and Multimedia Commission (MCMC) has simultaneously promoted digital expansion while intensifying online governance measures through takedown requests, licensing frameworks and content regulation. In 2024, it reported a significant rise in provocative and harmful online content, noting that 35,490 postings were removed in 2023, and a further 27,115 pieces of content within the first two months of 2024, including material relating to the so-called "3R" issues – the sensitive topics of race, religion and royalty – scams, hate speech, gambling and misinformation.^{26,27}

However, Malaysia's current governance trajectory risks reinforcing the problems identified above. Existing debates frequently focus on licensing regimes, takedown powers and stronger enforcement obligations. While harmful content remains a legitimate concern, control-oriented approaches fail to address the structural accountability problems embedded within platform architectures and business models. Regardless, the legislative framework governing digital communication and online harms remains fragmented and heavily oriented towards content regulation and enforcement.

The Communications and Multimedia Act (CMA) 1998 remains the primary legal framework regulating online communication, providing broad powers to the MCMC to address content deemed offensive, indecent, false, menacing, or harmful to public order. Effective 1 January 2025, the Malaysian government began enforcing a mandatory licensing regime for major social media and messaging platforms with more than eight million Malaysian users, justified as necessary to combat cybercrime, scams, cyberbullying and harmful content. The Online Safety Act 2025 (effective 2026) reflects growing state and public concern regarding harmful online content, cyberbullying, child safety, scams and digital harms circulating across platforms.

However, the tightening regulatory climate has been accompanied by growing concerns from civil society organisations about the continued use of the Sedition Act 1948 and Section 233 of the CMA against journalists, activists, politicians, artists and ordinary social media users. The latter, which criminalises the improper use of network facilities or services, has frequently been criticised

for its broad and vague wording, which enables selective enforcement against online criticism and dissent. High-profile cases include investigations against graphic artist Fahmi Reza's satirical online content, proceedings involving independent media portal Malaysiakini's apparently contemptuous reader comments, investigations into social media users accused of insulting royalty or religion and police probes against activists and politicians for online statements critical of government institutions.

Civil society actors argue that these legal provisions create a chilling effect on freedom of expression, particularly where online speech intersects with political criticism, minority rights or state accountability. Indeed, the Communications and Multimedia Content Forum of Malaysia (Content Forum) has documented increasing public anxiety over harmful online content, recording a sharp rise in complaints relating to internet and social media material (particularly involving cyberbullying, misinformation, offensive speech and inappropriate content on heavily used platforms).²⁸

These developments collectively demonstrate the contradictory dynamics within Malaysia's digital governance environment: one marked by high digital participation and connectivity, yet increasingly shaped by surveillance concerns, platform regulation, content moderation and the securitisation of online expression.

4. Deferring (digital) justice: social media, grievance and public accountability in Malaysia

The growing reliance on social media as an alternative grievance and accountability mechanism reflects broader structural tensions surrounding public governance systems. Digital platforms increasingly function as spaces where citizens seek visibility, recognition and institutional responses, while exposing deep weaknesses in formal accountability systems.

4.1 Declining trust despite effective formal mechanisms

One significant trend among citizens is the shift away from using formal grievance mechanisms towards relying on platform-based expression. Historically, complaints were channelled through government counters, written correspondence, telephone hotlines, tribunals or ombudsman-style mechanisms. But from the mid-2010s onwards, ministries and government agencies adopted digital systems, including websites, mobile applications and online portals. The widespread availability of messaging applications – which, in effect, turned government institutions themselves into social media users – saw their popular use for receiving public complaints.

Today, formal channels coexist with informal ones at the grassroots level, while chatbots are increasingly used to respond to queries by the private sector. Despite this expansion of media for reaching the government, public confidence in formal systems remains uneven. Citizens

frequently experience institutional complaint systems as fragmented and difficult to navigate. As some observers from the roundtable discussion noted, “current mechanisms are dead; we don’t know what they are or how to approach them”. Different agencies operate separate mechanisms with limited coordination or interoperability, sometimes competing to meet their respective key performance indicators. Complaint tracking systems may lack transparency, while responses are often delayed or unclear. As a result, many users perceive social media activity as being more effective at delivering responses, because it generates visibility and public pressure (as seen below).

There is evidence, however, that formal grievance mechanisms continue playing a substantial role in resolving disputes and delivering material outcomes. Systems such as the Public Complaints Bureau (Biro Pengaduan Awam), Public Service Complaint Management System (SiSPAA, Sistem Pengurusan Aduan Awam), sector-specific platforms such as e-Aduan (used across various ministries), Bank Negara’s complaints channel and the Tribunal for Consumer Claims Malaysia remain important institutional infrastructures. For example, as of 2025, this Tribunal received most of its complaints online (44,472), whether through WhatsApp or various official portals and applications (e-Aduan, EzAdu), compared with 82 letters, 1,070 face-to-face complaints and 80 telephone calls.²⁹ Two-thirds were resolved, with claimants successfully obtaining RM77.2 million in dues.³⁰

Thus, although public perceptions may favour informal or public (online) avenues, formal systems retain their practical effectiveness, particularly for claims requiring legal enforceability. The paradoxical coexistence of relatively effective formal mechanisms with declining public trust highlights a gap between institutional performance and user experience. Digitalisation alone has not resolved underlying issues of accessibility, responsiveness and procedural clarity. Many systems remain reactive rather than citizen-centric, lacking integrated feedback loops or real-time engagement.

4.2. Reshaping accountability ecosystems through platform logics

Overall, platformised governance reshapes Malaysian accountability ecosystems through three interrelated dynamics. *First*, social media platforms function as informal grievance and redress systems – in other words, they constitute a parallel enforcement mechanism – because public institutions are often perceived as inaccessible, fragmented, slow or ineffective. *Second*, platform architectures incentivise virality, outrage, emotional amplification and public shaming, creating conditions that reward spectacle rather than deliberation or procedural fairness. *Third*, increasing reliance on platform reporting systems and algorithmic moderation transfers governance away from democratic institutions towards private companies whose priorities are structured by commercial incentives rather than meeting public interest obligations. These dynamics will be studied in turn below.

Social media has been repurposed as an auxiliary layer of accountability, not necessarily as a replacement for formal channels, but a parallel mechanism used to accelerate responses, amplify grievances and impose reputational pressure on institutions. Virality arguably becomes a way of shifting the burden of responsibility (that is, reporting and following-up) from oneself to the masses, thus providing informal enforcement at scale. Indeed, complaints attracting widespread attention often trigger rapid institutional responses, such as seen with infrastructure.

Viral videos documenting dangerous roads, potholes or neglected public facilities frequently prompt immediate repairs after public outrage escalates online. Consequently, viral escalation is seen as capable of bypassing bureaucratic inertia. However, these successful responses reveal important asymmetries. Complaints only receive attention when they become visible enough to generate reputational pressure. A post on a Johor pothole (see Box 1), for instance, garnered 2.2 million views, but less visible grievances may remain ignored, regardless of severity. Justice therefore becomes contingent upon algorithmic amplification rather than procedural fairness.

Box 1: Fix that pothole!

Infrastructure complaints reveal how tagging and virality can produce rapid outcomes. A TikTok video of a man in Johor warning motorists about a dangerous pothole (“There is a large pothole. Please be careful and slow down”) triggered widespread engagement in March 2025. Users amplified the issue by tagging the multiple agencies responsible for road maintenance (for example, “@LLMtrafik @JKRMalaysia please fix this ASAP before someone dies” and “Already reported many times. @aduankl @KKRMalaysia no action!”). This strategy transformed a local issue into a national public accountability test: authorities responded immediately and repairs were completed shortly after the video spread widely.

Similar patterns appeared in other cases. A banana tree planted in a pothole went viral after being shared and local authorities were tagged, prompting repair work. A viral post by @MadaniNomics showed an elderly man repairing a pothole by himself, triggering public outrage against municipal councils. The contrast between the perceived inefficacy of formal complaints and the immediate actions resulting from tagging is evident from a sample user comment: “Report through apps nothing happens. Tag them here then suddenly move fast.”

Meanwhile, the combination of popular hashtags and tagging of official government social media accounts created sustained pressure in the wake of Zara Qairina Mahathir's death. Authorities responded by reopening investigations, conducting a second post-mortem and issuing updates. However, social media amplification also introduced distortions, where viral posts falsely implicated prominent figures or led to premature conclusions.

Box 2: #JusticeForZara

The death of Zara Qairina Mahathir demonstrates how tagging and hashtags can scale up grievance into a national movement. #JusticeForZara became a focal point for mobilisation, being widely shared across platforms. Users consistently tagged government authorities, ministers and agencies, as seen for example in three sample posts:

- "#JusticeForZara @PDRMsia please reopen investigation"
- "We demand answers @KemPendidikan @PMOMalaysia don't stay silent"
- "@AGCPutrajaya take action now!"

Offline mobilisation mirrored the online campaign. At rallies, participants held placards reading "Justice for Zara" and "Stop bullying", among others.

Similar dynamics were evident during eviction disputes in Kampung Sungai Baru, which was linked to redevelopment under the Land Acquisition Act. The incident gathered widespread online attention after videos of clashes and enforcement actions went viral. Residents and their supporters strategically used social media to frame redevelopment conflicts as broader questions about justice, displacement and political accountability. Tagging practices even extended beyond administrative agencies to implicate senior political leaders and royal institutions, thus appealing to a higher moral authority. Although digital mobilisation generated solidarity and national attention, it ultimately did not prevent the evictions.

Box 3: The Kampung Sungai Baru eviction

To fight eviction, residents and supporters used social media to escalate the issue through protest messaging which directly addressed political leaders ("PMX jangan berdiam @PMOMalaysia" [Prime Minister, do not remain silent]) and called for intervention from senior officials ("@anwaribrahim please intervene now. Stop this eviction"). Some users escalated the issue further by tagging royal institutions and accounts (for example, "@istananegara please look into Kampung Sungai Baru" and "@SelangorRoyalOffice rakyat need help"). This move reflects a broader pattern of going beyond administrative channels to appeal to increasingly symbolic or powerful actors otherwise distant from the issue.

Several residents and supporters framed the eviction as an injustice. "We were told to start packing up or police would break down the door", one resident claimed, and another individual called for direct results ("We want the government to intervene immediately... stop the evictions"). Despite high visibility, the eviction proceeded under court orders and debate shifted to broader issues such as urban renewal and property rights.

The politicisation of grievances further complicates the information environment. Certain issues are strategically amplified by coordinated networks or politically aligned actors, meaning that grievances may become intertwined with broader ideological mobilisation.³¹ Public accountability increasingly occurs within highly polarised and emotionally charged digital spaces, where so-called “injustice collectors”, despite claiming to represent public voices, often espouse narrow interests and can disproportionately influence overall sociopolitical narratives and discourses.

Religious disputes provide particularly important examples, where controversies involving temples, redevelopment or land ownership frequently escalate online into broader identity-based conflicts. Legal and administrative questions are reframed through symbolic narratives about discrimination, rights or state bias. Tagging practices directed at political leaders and ministries further intensify public pressure dynamics.

While authorities respond with statements or temporary interventions, as seen in the case of Hindu temples built on land with historical or informal permission (yet lacking the necessary legal documents), the underlying challenge related to land governance remains unresolved. The disproportionate responses to this issue lie in how legal and technical questions (for example, around land status, planning approval and redevelopment relocation negotiations) become more fundamental and symbolic conflicts about religious rights and state bias through online amplification.

Box 4: Hindu temple disputes

Disputes over Hindu temples are frequently politicised once they enter social media circulation. Although the problems are rooted in legal and technical considerations, they are quickly amplified into identity-based conflicts online. Some sample posts demonstrate how users signal the parties responsible or indicate who should act, by mobilising wider audiences and making the issue visible to influential accounts. For example:

- “Why demolish a temple that is more than 100 years old? @PMOMalaysia must answer.”
- “Sounds fakey... why demolish a temple to build a mosque? @DBKL2u please explain.”
- “When it comes to temples, everyone becomes scared. @JKRMalaysia do something.”

In some threads, users also tagged multiple agencies simultaneously (for example, @PMO, @DBKL2u, @KPN or @MalaysiaPM) to create a sense of urgency and pressure.

Overall, the relationship between amplification and resolution remains highly uneven. Social media excels at creating exposure, but performs poorly as a sustained infrastructure of accountability. In many cases, a combination of exposures and mass media coverage (bearing in mind that both forms are increasingly intertwined) focuses attention on key exchanges, highlights actors involved and results in intermediate actions (such as protests and police investigations), but rarely offer

resolution. Discussions often fragment along competing narratives, while attention rapidly shifts towards newer controversies. Formal systems at least provide continuity, documentation and procedural pathways, even if only in theory, in contrast to social media's privileging of immediacy and emotional momentum.

In this context, social media's affordance of anonymity is a key factor in its selection as a recourse. Anonymity can be necessary, especially when the risks are high for those seeking remedies to openly speak out, yet it can also facilitate harms against others. In some cases, it provides some form of accessibility and emotional validation absent from formal systems, particularly for individuals who may feel safer disclosing their experiences online, due to perceived anonymity or the ability to garner collective support. This is particularly so in contexts shaped by hierarchical power relations or fears of retaliation, especially among those who are marginalised or made vulnerable. At the same time, anonymity complicates accountability and verification. False allegations and misinformation can spread rapidly, while hostile actions, such as coordinated harassment or doxing, are facilitated. Corrections rarely receive equivalent visibility once initial claims circulate widely, thus creating enduring reputational harms even after allegations are later proved inaccurate.

Overall, algorithmic amplification intensifies these dynamics. Since platforms reward emotionally charged content, users therefore adapt their communication strategies accordingly, crafting posts designed to maximise shareability and emotional responses. Visual content, provocative framing and targeted tagging become core digital grievance strategies. These developments reveal important governance gaps, where profit-driven social media platforms increasingly usurp the role of democratically accountable formal pathways and institutions. Public institutions increasingly operate reactively to viral pressure rather than through proactive accountability mechanisms.

In effect, Malaysians are experiencing a broader digital displacement, rather than moving towards genuine digital justice. Accountability functions are gradually shifting away from public (albeit bureaucratic and sometimes slow) institutions towards opaque platform systems governed by commercial incentives and engagement metrics. Here, justice is increasingly conceptualised in terms of and reduced to, online visibility, in line with unaccountable platform logics.

4.3. Policy gaps and governance challenges

Malaysia's digital governance landscape is characterised by fragmentation, reactive enforcement and insufficient attention to systemic accountability. Existing approaches frequently prioritise content control and reputational management, while neglecting the structural dynamics shaping digital participation.

One major gap is the absence of *integrated accountability systems*. Public complaint mechanisms remain dispersed across ministries, agencies, local authorities and sectoral regulators. Citizens

often struggle to identify appropriate channels, track complaints or obtain meaningful feedback. In practice, social media visibility increasingly functions as an informal escalation pathway to compensate for institutional fragmentation.

Another significant weakness is *limited platform transparency obligations*. Technology companies operating within Malaysia face relatively few requirements relating to algorithmic accountability, moderation transparency or public disclosure in response to government requests. As a result, citizens possess limited understanding of how visibility systems shape discourse or how moderation decisions are implemented.

Current governance approaches also *inadequately address algorithmic amplification systems*. Regulatory attention remains heavily focused on individual pieces of content – that is, content moderation – rather than the underlying recommendation architectures and engagement incentives driving harmful dynamics. Yet evidence increasingly demonstrates that platform harms emerge systemically through amplification logics, rather than solely through isolated unlawful content from individuals.

Procedural safeguards also remain weak. Users encountering content removal, harassment, misinformation or reputational harm frequently lack accessible remedy pathways. Platform reporting systems remain inconsistent while legal remedies may be inaccessible, expensive or excessively slow.

Additionally, there remains *limited digital governance expertise at an institutional level*. Effective oversight increasingly requires a technical understanding of recommendation systems, data infrastructures, automated moderation tools and platform business models. However, public institutions often lack sufficient interdisciplinary expertise to engage these issues comprehensively. Perhaps most importantly, *current governance trajectories risk conflating safety with control*. Expansive regulatory powers introduced without strong safeguards may produce chilling effects on freedom of expression and political participation, thus running counter to democratic norms. Vulnerable communities, activists and independent journalists may experience disproportionate risks under opaque enforcement systems ostensibly designed to improve public safety.

5. Policy recommendations

Ultimately, Malaysia's digital future should not be governed primarily by the logic of virality, surveillance or engagement maximisation, which feed into unchecked profit-making or centralised political power. A democratic digital environment requires transparent institutions, accountable platforms, strong public safeguards and meaningful access to remedy. (Digital) justice cannot depend solely on visibility, but must be grounded in rights, fairness and accountability. Taken together, these recommendations recognise that online harms and digital injustices cannot be addressed only through punitive regulation or platform self-governance. Sustainable reform requires coordinated action from government institutions, legislators, technology companies, civil society, researchers and the public (that is, as digital citizens). The following recommendations are therefore targeted towards the key stakeholders responsible for specific reforms and are intended to support actionable policy development.

In general, Malaysia requires a rights-based and institution-centred digital governance framework that strengthens democratic accountability without expanding arbitrary state or corporate influence over public expression. Its policy response must therefore move beyond reactive content control and instead towards systemic governance that protects fundamental freedoms and restores public trust in both digital platforms and state institutions alike.

5.1. Government and regulatory agencies

5.1.1. Strengthen public grievance and complaint systems

Federal and state agencies should modernise and integrate existing complaint systems (such as SiSPAA, e-Aduan, consumer tribunals and municipal complaint platforms), where possible. This would encourage a more accessible and effective process that ensures adequate follow-through and is integrated with other accountability systems (for example, civil society monitoring, media scrutiny and judicial processes).³² In this case, government agencies should:

- Introduce interoperable complaint-tracking systems;
- Provide mandatory response timelines;
- Publish meaningful complaint resolution statistics;
- Improve accessibility for persons with disabilities, rural communities and disadvantaged communities; and
- Provide multilingual complaint mechanisms.

5.1.2 Institutionalise feedback and enforceable remedies

Grievance systems should ensure that complaints translate into outcomes that are visible and enforceable, to address public dissatisfaction over the perceived lack of resolutions for their complaints. Agencies should build the following reforms into their processes:

- Provide structured feedback at each stage of the complaint process, including reasons for decisions and timelines for next steps;
- Establish automatic escalation pathways for unresolved or delayed cases within and across agencies; and
- Link complaint outcomes to corrective actions, by escalating complaints to independent review and institutional misconduct mechanisms.

5.1.3 Integrate online-offline support

While digitisation has expanded access, an over-reliance on online systems can exclude users who require assistance and limit the type of grievances that can be resolved. Hybrid options that allow for face-to-face interactions are particularly important for those with limitations. To encourage the use of grievance mechanisms, agencies should:

- Ensure adequate people-centred and people-run response systems to balance out automated functions;
- Develop structured resolution units to handle complex or sensitive complaints that cannot be easily addressed by existing work processes; and
- Promote opportunities for dialogue, clarification of facts and negotiation of outcomes.

5.1.4 Reform laws to establish an online safety framework

The Communications Ministry and the MCMC should review the CMA, particularly Section 233, to ensure that enforcement powers comply with constitutional protections for freedom of expression and international human rights standards. Definitions relating to offensive, false or harmful content should be narrowed and clarified to prevent arbitrary or selective enforcement. Meanwhile, the implementation of the Online Safety Act should prioritise systemic accountability and procedural safeguards rather than expand broad censorship powers. Regulations should distinguish between illegal content, lawful but harmful speech and political expression to avoid collateral censorship. The legal framework should also be robust enough to address legitimate limitations, such as the need to protect the most vulnerable users.

5.1.5. Introduce mandatory platform transparency obligations

Malaysia should establish statutory transparency obligations, especially for the large licensed service providers (with over eight million registered users) operating in the country. The licensing requirement came into effect on 1 January 2026, and is provided for in the CMA's Class Licence for Application Service Providers. Platforms should be required to publish:

- Government takedown requests and legal demands;
- Data on moderation decisions and appeals outcomes;
- Disclosure of automated moderation systems usage;
- Advertising information (political and issue-based advertising); and
- Data on coordinated disinformation and manipulation campaigns.

The MCMC should also publish its own periodic transparency reports detailing all content restriction requests issued by state agencies, ministries and law enforcement bodies.

5.1.6. Establish independent audits of recommender systems and algorithmic amplification mechanisms

Drawing from models such as the European Union's DSA, Malaysia should require independent third-party audits of recommender systems and algorithmic amplification mechanisms. Audits should be conducted by accredited independent researchers or institutions, rather than by platforms themselves. They should assess risks relating to:

- Amplification of hate speech and religious polarisation;
- Gender-based violence and coordinated harassment;
- Child safety and cyberbullying;
- Electoral manipulation and political disinformation; and
- Discriminatory moderation practices.

5.1.7. Create an independent oversight and governance commission

Malaysia should establish an independent multi-stakeholder digital rights and platformised governance commission to enable oversight and governance separate from direct ministerial control. This commission should include civil society organisations, technical and cybersecurity experts, human rights advocates, media practitioners, academics and representatives from vulnerable and marginalised communities. Its mandate should include:

- Reviewing platformised governance policies;
- Monitoring government requests for content removal;
- Advising Parliament on digital governance reforms;
- Conducting public consultations; and
- Receiving complaints relating to platformised governance and digital rights.

5.2. Parliamentary representatives and policymakers

5.2.1. Adopt a rights-based digital governance framework

Parliament should adopt a comprehensive digital governance framework grounded in constitutional rights, democratic accountability and international human rights standards, including Article 19 of the UDHR on freedom of expression and the UN's Guiding Principles on Business and Human Rights. Digital governance legislation should incorporate:

- Necessity and proportionality safeguards;
- Judicial oversight for content restrictions;
- Independent review and appeals mechanisms;
- Public reporting obligations; and
- Protections against politically motivated enforcement.

5.2.2. Strengthen oversight of digital regulation

A cross-party select committee on digital governance and online harms should be established to provide continuous oversight of platform regulation, government enforcement and emerging digital risks. This committee can conduct public hearings with platforms and regulators, review annual transparency reports, conduct consultations with civil society and researchers as well as evaluate the impact of online safety laws on civic participation and media freedom.

5.2.3. Develop national standards on platform accountability

Policymakers should develop national standards on platform accountability through multi-stakeholder consultations to address, among other matters, transparency reporting, appeals and remedy systems, data governance, privacy protection, political advertising and platform responses to coordinated harms. Malaysia should align these standards with emerging global best practices (such as outlined in Section 3.1) while adapting them to local democratic and multicultural realities.

5.3. Private sector and digital platforms

5.3.1. Improve content moderation transparency and appeals processes

Platforms operating in Malaysia should improve procedural fairness and transparency in moderation practices, beginning with the use of third-party providers of moderation services, their own automated moderation systems and overall artificial intelligence (AI) training practices.³³ Platforms should also provide meaningful explanations for removals and restrictions, offer accessible appeals mechanisms in Bahasa Malaysia and other local languages, improve responsiveness to harassment and abuse reports as well as reduce an over-reliance on AI moderators. Special attention should be given to the protection of groups such as journalists, women, activists, minority communities and human rights defenders, who experience disproportionate online harms.

5.3.2. Conduct human rights impact assessments

Technology companies should conduct regular human rights impact assessments on how their recommendation systems, advertising models and moderation practices affect Malaysian users.³⁴ These assessments could examine the use and impact of content related to religious and racial polarisation, harassment, misogyny, political manipulation and threats to freedom of expression, among others. Importantly, these findings should be made publicly available.

5.3.3. Support independent research and data access

Platforms should provide secure and privacy-respecting data access mechanisms for independent researchers, universities and civil society organisations studying online harms, algorithmic amplification and digital participation. Limited access to platform data currently prevents meaningful public accountability and evidence-based policymaking.

5.3.4. Protect user privacy and political participation

Identity verification systems linked to centralised databases or mandatory real name registration create significant surveillance risks and may discourage political participation and online expression. Therefore, platforms and regulators should instead adopt privacy-enhancing technologies and decentralised verification systems that protect anonymity while addressing legitimate safety concerns.

5.4. Civil society, academic and media actors

5.4.1. Expand independent digital rights monitoring

Independent monitoring is essential to counter opaque governance practices and strengthen public accountability. Civil society organisations, media watchdogs and academic institutions should establish long-term monitoring programmes examining:

- Government censorship requests;
- Platform moderation practices;
- Coordinated disinformation campaigns;
- Online gender-based violence; and
- Algorithmic amplification of harmful content.

5.4.2. Build strategic coalitions across sectors

Civil society organisations should strengthen collaborations with journalists, legal experts, consumer associations, youth groups and community organisations, among others, to advocate for rights-based digital governance. Universities and think-tanks could offer collaborative opportunities for interdisciplinary research on platformised governance, online harms, algorithmic systems and democratic participation. Such coalition-building is necessary to ensure that online safety debates do not become solely dominated by state security actors or corporate interests.

5.5. The general public/digital citizens

5.5.1. Strengthen public digital literacy beyond “fake news” campaigns

Digital literacy programmes should be integrated into schools, universities and community education programmes. They should move beyond simplistic anti-misinformation and anti-disinformation campaigns to broaden public understanding of:

- Algorithmic amplification;
- Platform business models;
- Privacy and surveillance risks;
- Online harassment dynamics; and
- Digital rights and access to remedy.

At the same time, political parties and their agents also play an important role in preventing mobilisation around and politicisation of unresolved issues or public grievances. As intermediaries between institutions and citizens, they should support literacy efforts among party members and constituents on meaningful participation and accountability in a platform-mediated public sphere.

5.5.2. Promote responsible digital participation

Campaigns should encourage responsible online participation among the public, including:

- Verifying factuality before sharing allegations;
- Responsibly tagging accounts and amplifying narratives;
- Protecting oneself against doxing and online vigilantism; and
- Maintaining respect for due process, dignity and privacy.

Users also need to understand the relationship between the information and business architectures of platform technologies, their desires to express views or grievances as well as the potential benefits and harms of digital activity.

5.5.3. Improved public engagement and education

Digital governance should not be determined solely by negotiations between governments and technology companies. Meaningful public participation is necessary to ensure legitimacy, accountability and trust. Therefore, digital citizens should actively engage in consultations relating to online safety laws, platform regulation and digital rights reforms.

6. Conclusion

Malaysia stands at an important crossroads in the evolution of digital governance and pursuit of digital justice. Social media platforms increasingly mediate public accountability, political participation and access to visibility, with citizens turning to digital platforms to gain recognition, pressure institutions as well as mobilise support and responses. These practices also reveal deep structural weaknesses within both public governance systems and platform infrastructures. Current governmental approaches focus heavily on granular reactive measures (for example, takedowns and licensing), even as online harms increasingly systemic (shaped by platform architectures, algorithmic amplification, opaque moderation systems), while unequal access to remedy persists for particular groups.

In a strong democratic system, citizens *should not* have to rely on social media virality, mass outrage or public humiliation simply to obtain digital justice (even if these platforms are useful for lowering barriers to accessing formal mechanisms for some). The main challenge is not whether digital platforms should play a role in accountability ecosystems. They already do. Rather, the challenge lies in determining the emerging order of accountability and the interests that it ultimately serves. Current trends and practices in Malaysia, as elsewhere, risk producing systems where justice becomes dependent on virality, emotional amplification and algorithmic visibility. Such systems may provide symbolic recognition or temporary pressure, but hardly guarantee fairness, procedural integrity or structural reform. Public institutions respond reactively to digital outrage by centralising authority, while private corporations continue exercising authority without equivalent accountability obligations.

Ultimately, *justice cannot be reduced to visibility alone*. A democratic digital society requires institutions and systems capable of delivering accountability beyond the logic of virality, in order to prevent further digital displacement. We need to move beyond narrow content regulation frameworks and towards systemic governance approaches grounded in rights, transparency, accountability and oversight. Platformised governance must be subjected to meaningful scrutiny while public institutions must rebuild legitimacy through accessibility, responsiveness and procedural fairness. The future of Malaysian digital governance should not be defined solely by state control or corporate self-regulation. Instead, it requires democratic infrastructures capable of protecting vulnerable communities, preserving political agency and ensuring meaningful access to remedy.

Endnotes

¹ Gillespie, T. (2018). *Custodians of the internet*. Yale University Press.

² Suzor, N. (2019). *Lawless: The secret rules that govern our digital lives*. Cambridge University Press.

³ Liu, D. (2024). Borderline content and platformised speech governance: Mapping TikTok's moderation controversies in South and Southeast Asia. *Policy & Internet*, 16(3), 543–66. <https://doi.org/10.1002/poi3.388>

⁴ Legocki, K.V., Walker, K.L. & Kiesler, T. (2020). Sound and fury: Digital vigilantism as a form of consumer voice. *Journal of Public Policy & Marketing*, 39(2), 169–87.

⁵ Venkiteswaran, G. (2017). "Let the mob do the job": How proponents of hatred are threatening freedom of expression and religion online in Asia. Association for Progressive Communications. <https://www.apc.org/en/pubs/let-mob-do-job-how-proponents-hatred-are-threatening-freedom-expression-and-religion-online>

⁶ Loveluck, B. (2020). The many shades of digital vigilantism. A typology of online self-justice. *Global Crime*, 21(3–4), 213–41. <https://www.tandfonline.com/doi/full/10.1080/17440572.2019.1614444>

⁷ Cox, C.A.R. (2021). Taking the Law into their (Virtual) Hands: Facebook and digital vigilantism. *Archium. ATENEO*. <https://archium.ateneo.edu/comm-faculty-pubs/29>

⁸ Ibid., p. 2.

⁹ Dannhauser, L.M. (2024). Protecting the Innocent: How to Prevent the Consequences of Misidentification and Doxing by Volunteers Helping with Open Source Investigations. *Catholic University Journal of Law and Technology*, 32(2). <https://scholarship.law.edu/jlt/vol32/iss2/3>

¹⁰ Celeste, E., et al. (2023). *The content governance dilemma*. Palgrave Macmillan.

¹¹ OHCHR (2018, April 6). Report on content regulation issued by Special Rapporteur on freedom of opinion and expression (Document A/HRC/38/35). <https://www.ohchr.org/en/calls-for-input/report-content-regulation>

¹² Council of Europe. (n.d.). Digital Citizenship Education. [https://www.coe.int/en/web/education/digital-citizenship-education#\(%22271421625%22:\[2\]\)](https://www.coe.int/en/web/education/digital-citizenship-education#(%22271421625%22:[2]))

¹³ OHCHR. (2025). Threats to freedom of expression online in turbulent times (Document A/80/341). <https://docs.un.org/en/A/80/341>

¹⁴ Santa Clara Principles. (n.d.). Santa Clara Principles on Transparency and Accountability in Content Moderation. <https://santaclaraprinciples.org/>

¹⁵ European Union. (n.d.). European Union Digital Services Act 2022/2065. <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act>

¹⁶ Government of Australia. (2021). Online Safety Act 2021. <http://legislation.gov.au/C2021A00076/latest/text>

¹⁷ Government of the UK. (2023). Online Safety Act 2023. <https://www.legislation.gov.uk/ukpga/2023/50>

¹⁸ Luck, P., & Gray, R. (2026, Feb 3). Where sovereignty meets statecraft: South Korea's Fairness Act and the U.S. Alliance.

Centre for Strategic and International Studies. <https://www.csis.org/analysis/where-sovereignty-meets-statecraft-south-koreas-fairness-act-and-us-alliance>

¹⁹ Examples include YouTube's Harassment Policy and Facebook/Meta's Transparency Report.

²⁰ Baccarella, C., et al. (2020). Social media? It is serious! Understanding the dark side of social media. *European Management Journal*, 36(4), 431–38. <https://doi.org/10.1016/j.emj.2018.07.002>

²¹ ARTICLE 19. (2021). Social media councils: One piece in the puzzle of content moderation. URL: <https://www.article19.org/wp-content/uploads/2021/10/A19-SMC.pdf>; PEN America. (2023). Shouting to the void: Why reporting abuse to social media platforms is so hard and how to fix it. <https://pen.org/report/shouting-into-the-void/>; Ranking Digital Rights. (2025). 2025 RDR index: Big Tech Edition. <https://rankingdigitalrights.org/bte25/>

²² Diaz, A. & Hecht-Felella, L. (2021). Double standards in social media content moderation. Brennan Center for Justice at New York University School of Law. <https://www.brennancenter.org/our-work/research-reports/double-standards-social-media-content-moderation>

²³ Celeste (2023).

²⁴ PEN America (2023).

²⁵ Kemp, S. (2025). Digital 2025: Malaysia. DataReportal. <https://datareportal.com/reports/digital-2025-malaysia>

²⁶ Bernama. (2024, April 9). MCMC: Significant rise in provocative content on social media. <https://focusmalaysia.my/mcmc-significant-rise-in-provocative-content-on-social-media/>; Welsh, B. (2020, 18 August). Malaysia's Political Polarization: Race, Religion, and Reform. Bridget Welsh. <https://bridgetwelsh.com/articles/malaysias-political-polarization-race-religion-and-reform/>; Yang Razali Kassim. (2020, 2 December). Malaysia' Political Crisis: A new power twist? (RSIS Commentary). S. Rajaratnam School of International Studies. <https://rsis.edu.sg/wp-content/uploads/2020/12/CO20206.pdf>.

²⁷ The prevalence of identity politics in the everyday lives of Malaysians has been long established, particularly around ethnicity and religion. These, together with the position of the monarchy, have been subjects of tensions and controversies over the years. However, it was only in the 2010s (especially after the 14th general election in 2018) that the term "3Rs" entered mainstream discourse. Post-2018 political contestations have marked unprecedented royal intervention into post-independence politics.

²⁸ Communications and Multimedia Content Forum (2023, January 31). 71% of Complaints to the Content Forum are on Social Media Content (press release). <https://contentforum.my/wp-content/uploads/2023/08/71-percent-of-Complaints-to-the-Content-Forum-are-on-Social-Media-Content.docx.pdf>

²⁹ Domestic Trade and Cost of Living Ministry. (n.d.). Statistik Utama KPDN 2025. <https://www.kpdn.gov.my/ms/media-utama/penerbitan/statistik-utama-kpdn>

³⁰ Farah Nabilah Wazer. (2025). Consumer Tribunal resolves nearly 7,000 cases involving RM77mil in claims. *New Straits Times*. <https://www.nst.com.my/news/nation/2025/11/1320676/consumer-tribunal-resolves-nearly-7000-cases-involving-rm77mil-claims>

³¹ Danish Raja Reza. (2025, 25 November). How the Zara Qairina tragedy became caught in political crossfire. *Free Malaysia Today*. <https://www.freemalaysiatoday.com/category/nation/2025/11/25/how-the-zara-qairina-tragedy-became-caught-in-political-crossfire>

³² Pande, S., & Hossain, N. (2022). Grievance redress mechanisms in the public sector: A literature review. Accountability Research Centre and the Open Government Partnership. <https://www.opengovpartnership.org/wp-content/uploads/2022/05/Grievance-Redress-Mechanisms-in-the-Public-Sector-A-Literature-Review.pdf>

³³ Services are provided by third-party companies to clean up and train platform AI, while systems are the platform's built-in infrastructure. For example, training and refining a large language model could be a service provided to TikTok, where automation (instead of human moderation) is the system adopted by the platform.

³⁴ Venkiteswaran, G., Lim, S., & Naidu, W. (2023). Voting & Shopping: evaluating media & technology companies in Malaysia in the context of elections & e-commerce. Digital Asia Hub. <https://cijmalaysia.net/voting-shopping-evaluating-media-technology-companies-in-malaysia-in-the-context-of-elections-e-commerce/>

