

# State, Morality, and the Digital Public Sphere: A Comparative Study of Content Moderation Rationales in China and Pakistan

Iqra Riaz\*, Nouman Ali

Sargodha, Pakistan

\*Corresponding Author

DOI: <https://doi.org/10.47772/IJRISS.2026.100701026>

Received: 02 August 2026; Accepted: 07 August 2026; Published: 19 August 2026

## ABSTRACT:

Content moderation regimes are typically studied within their own national and political contexts, and rarely compared across ostensibly incompatible political systems. This paper compares the moral rationale of content moderation in China and Pakistan, two states that differ sharply in regime type, religious establishment, and legal tradition, to examine whether a common structural pattern underlies their otherwise distinct architectures. Drawing on China's Cyberspace Administration "Qinglang" campaigns and Pakistan's Prevention of Electronic Crimes Act (PECA) of 2016 and its 2025 amendment, the analysis compares three axes of moderation: the definitional vocabulary used to justify harm, the institutional authority empowered to adjudicate it, and the operational logic through which enforcement proceeds. Existing scholarship has treated Chinese cyber-sovereignty and Pakistani blasphemy-obscenity regulation as largely separate literatures, leaving open what, if anything, connects authoritarian and hybrid-democratic content regulation. This paper addresses that gap through a documentary comparative analysis of primary regulatory texts, government and non-governmental reports, and peer-reviewed scholarship, structured through a most-different systems design. The analysis suggests that despite divergent definitional grammars and institutional architectures, both regimes are consistent with a recurring pattern of regulatory and extra-legal attention to women's digital visibility. This pattern is conceptualized as digital pardah: a reconstruction of modesty governance in which platform infrastructure, statutory language, and informal social authority jointly, rather than singly, produce the conditions under which women's online presence becomes governable. The paper argues that digital pardah captures a mechanism, the joint and often uncoordinated action of state, platform, and family authority on the same body, that existing frameworks of platform governance, algorithmic visibility, and digital authoritarianism do not individually specify.

**Keywords:** content moderation, digital pardah, gender, platform governance, digital authoritarianism

## INTRODUCTION

The regulation of online content has become one of the central sites through which contemporary states assert moral and political authority. Where earlier sociological accounts of modesty, propriety, and public morality were anchored in family, mosque, temple, or neighborhood (Mahmood, 2005; Abu-Lughod, 1986), a significant share of that regulatory work has migrated to platforms, algorithms, and state cyber-agencies. This paper asks a comparative question: when two states with markedly different political systems and religious-cultural orientations both undertake large-scale content moderation, what moral grammar do they use to define harm, who is authorized to adjudicate it, and how is enforcement carried out?

China and Pakistan present an instructive, most-different comparison within the broader field of digital authoritarianism studies (Roberts & Oosterom, 2024; Feldstein, 2021). China is a one-party state with an elaborate, centralized cyber-governance apparatus; Pakistan is a hybrid electoral democracy in which digital regulation is shaped by an uneasy coalition of state agencies, courts, and religious-political actors. Both states regulate content in the name of protecting a collective moral order, one framed around social stability and

national values, the other around religious sanctity and cultural authenticity, and both, as the analysis below suggests, converge in treating women's online visibility as a particular object of concern.

This paper is deliberately narrow in scope. Rather than surveying the full architecture of platform law in each country, it focuses on a single analytical problem: the moral rationale for content moderation and its gendered consequences. This focus sits within the sociology of morality, gender, and media rather than within comparative law, and builds on existing scholarship examining how modesty norms are reconstructed, rather than simply transposed, in digital space. Section 2 reviews and synthesizes the relevant literature and states the resulting research gap explicitly. Section 3 develops the comparative methodology and the conceptual contribution of digital *pardah*. Section 4 applies the framework to China and Pakistan in turn. Section 5 examines the gendered convergence across both cases. Section 6 concludes with theoretical, methodological, and policy implications.

## LITERATURE REVIEW

### Platform Governance and the Politics of Visibility

Sociological accounts of morality have long treated the designation of certain acts or images as harmful or indecent as a form of boundary-work through which collectivities define and defend their moral order (Durkheim, 1893/1984). Foucault's (1977) account of governmentality extends this insight to modern surveillance, arguing that power operates not only through prohibition but through the continuous, often invisible, structuring of visibility and conduct. Content moderation scholarship has applied this lens to digital platforms, describing moderation systems as a form of custodianship that quietly constitutes the public sphere even as it claims merely to clean it up (Gillespie, 2018; Roberts, 2019; Gorwa et al., 2020). This literature is largely built around a single institutional actor, the commercial platform, adjudicating what to remove.

A parallel and only partially reconciled literature on algorithmic visibility complicates this picture. Bucher (2012) argues that the more consequential mechanism of platform power is not removal but differential visibility: users experience a persistent threat of invisibility, of being algorithmically deprioritized rather than deleted, that disciplines behavior more continuously than outright bans. The two literatures disagree, implicitly, about where the center of gravity of platform power lies. Moderation scholarship treats the moment of removal as the analytically decisive event; visibility scholarship treats the ongoing, graduated distribution of attention as decisive, with removal as only its most extreme instance. The cases examined in this paper sit uneasily between the two: state-directed content governance in China and Pakistan operates chiefly through explicit removal and blocking, the classical moderation scenario, yet the pattern this paper identifies, that women's accounts and content recur disproportionately across otherwise unrelated moderation categories, resembles the selective, differential logic that visibility scholarship describes for commercial engagement metrics rather than state morality categories. This unresolved tension between removal-centered and visibility-centered accounts of platform power is one of the debates this paper's comparison speaks to.

### Digital Authoritarianism: A Fractured Field

Scholarship on Chinese internet governance has emphasized cyber-sovereignty and the state's preoccupation with stability maintenance, in which online expression is regulated less as individual morality than as collective social order (Creemers, 2017; Han, 2016; MacKinnon, 2011). Feldstein (2021) extends this into a broader comparative typology of digital repression, identifying surveillance, censorship, social manipulation, cyberattacks, internet shutdowns, and targeted persecution as recurring techniques across authoritarian and hybrid regimes alike. His framework is valuable precisely because it is built for cross-regime comparison rather than single-country description, yet the empirical base of this literature remains weighted toward consolidated authoritarian states and toward comparisons among them. Pakistan, a hybrid electoral democracy in which digital regulation is shaped as much by religious-political contestation as by state administrative capacity, sits awkwardly within a literature built primarily around consolidated one-party or personalist rule. Scholarship specific to Pakistan (Jamil, 2021; Ahmed et al., 2024) documents the country's own trajectory toward digital authoritarian practices but has developed largely independently of the China-centered literature above, with little cross-citation in either direction.

The recurring Qinglang campaigns coordinated by the Cyberspace Administration of China illustrate the collectivist orientation of the Chinese literature: since 2016, and on an annual cycle from 2020 onward, these campaigns have targeted pornography and vulgarity, cyberbullying, celebrity fan-circle conflict, the flaunting of wealth, and, more recently, generative artificial intelligence misuse and the incitement of gender antagonism (Wang et al., 2025; Xinhua, 2025; Global Times, 2025, 2026). Pakistan's Prevention of Electronic Crimes Act (PECA) of 2016 instead fuses religious, sexual, and political categories of harm within a single statutory clause authorizing removal of content that offends the glory of Islam, threatens state security, or violates public order, decency, or morality (Government of Pakistan, 2016; Al Jazeera, 2020). Analysts have repeatedly noted that this conflation leaves considerable discretion to whichever actor, regulator, court, or complainant, first invokes it (Global Voices, 2020; TechPolicy.Press, 2024).

### **Gendered Content Moderation: Convergent Outcomes, Divergent Mechanisms**

A third body of literature bridges these regulatory traditions through gender, though it too contains unresolved disagreement about mechanism rather than outcome. Banet-Weiser (2018) argues that popular feminism and popular misogyny are mutually constitutive rather than opposed: platforms reward feminism as marketable, brand-safe content while simultaneously treating feminist speech as politically disruptive, producing structurally contradictory treatment of the same content depending on context. Liao (2024) pushes further, arguing that China's state-market nexus does not merely tolerate misogynist content but actively platformizes it because it serves commercial engagement, a claim in some tension with the official rationale, expressed through categories such as gender antagonism, that campaigns like Qinglang exist to protect social harmony from gendered conflict. If Liao's account is correct, the enforcement priority nominally protecting women from antagonism may in practice fall more heavily on feminist speech than on the misogynist content that motivated it, a pattern documented directly in the removal of the Feminist Voices Weibo account in 2018 (Rest of World, 2023) and in the treatment of censored feminist accounts as a form of enforced digital disappearance (Shao & He, 2024).

A further complication comes from Wu and Dong (2019), who document that Chinese feminism is itself internally fragmented along class lines, with urban, educated, and internationally networked feminist voices occupying a different position, both discursively and in relation to state scrutiny, than working-class or rural articulations of gender grievance. This matters because it means state targeting of feminist accounts may be selectively engaging one class fraction of a heterogeneous feminist field rather than feminism uniformly, complicating any simple narrative of state versus feminists. On the Pakistani side, Naseer and Ashraf (2022) apply the CEDAW human-rights framework to argue that the state's own weak internet-security mechanisms, not only its moral categories, contribute to gendered digital harm, since inadequate platform-level protections leave women's accounts more exposed to coordinated harassment regardless of the content they post. Digital Rights Foundation documentation is consistent with this account, showing women as a majority of reported harassment victims and describing particular pressure on women journalists and public figures toward self-censorship (Digital Rights Foundation, 2024, 2026). These two national literatures, Chinese feminist media studies and Pakistani digital-rights and legal scholarship, rarely cite one another despite describing what this paper treats as a structurally comparable phenomenon.

### **The Research Gap**

Three gaps follow from this review. First, digital authoritarianism scholarship has developed largely through comparison among consolidated authoritarian states or through single-country case studies, leaving underexamined how a hybrid electoral democracy with an explicitly religious legal register compares to a one-party, explicitly secular-collectivist system. Second, gendered-moderation scholarship has documented gendered effects within each national context largely in isolation: work on Chinese feminist censorship (Liao, 2024; Shao & He, 2024; Wu & Dong, 2019) and work on Pakistani cyber-harassment and blasphemy-obscenity law (Naseer & Ashraf, 2022; Jamil, 2021) rarely engage each other's cases, despite describing what appears to be a structurally similar phenomenon. Third, no existing framework specifies the particular mechanism by which modesty, rather than visibility, harm, or patriarchy in general, becomes reconstructed as a governable digital object jointly produced by state, platform, and family authority acting without formal coordination among them. Comparing China and Pakistan through a most-different systems design addresses the first two gaps by testing whether a common pattern survives near-total variation in regime type and legal tradition. The concept of digital

pardah, developed in Section 3.4, addresses the third gap by naming a mechanism that existing vocabularies, digital surveillance, platform governance, visibility politics, algorithmic governance, digital modesty, gendered moderation, platform patriarchy, and governmentality, do not individually capture.

## CONCEPTUAL FRAMEWORK AND METHODOLOGY

### Comparative Design and Case Selection

This paper adopts a most-different systems design (Przeworski & Teune, 1970), comparing two states that diverge sharply in regime type, religious establishment, legal tradition, media ownership structure, and economic system. China and Pakistan were selected because each offers a single, well-documented regulatory anchor suitable for sustained comparison, the Qinglang campaign cycle and the Prevention of Electronic Crimes Act respectively, and because both states justify large-scale content regulation through an explicit vocabulary of collective morality rather than through purely technical or commercial rationale. A most-different design is appropriate here, rather than a most-similar design pairing two authoritarian states or two Muslim-majority states, because the analytical interest is in whether a pattern survives variation in every institutional variable ordinarily thought to explain content regulation. If a convergence appears despite this variation, it is harder to attribute to shared regime type, shared religious establishment, or shared legal tradition, and correspondingly more plausible as a structural feature of content governance itself.

The design has a clear limitation, which is addressed directly in Section 3.5: a two-case most-different design maximizes analytical leverage over the reach of a convergence but does not, by itself, establish how common the pattern is across the wider population of content-moderating states. The claims that follow are accordingly framed as evidence that a pattern can survive extreme institutional variation, not as an estimate of its prevalence worldwide.

### Data Corpus and Document Selection

The corpus draws on four document types. Primary regulatory texts include the PECA 2016 statute and its 2025 amendment, and Cyberspace Administration of China campaign announcements as reported through Xinhua and People's Daily Online, since the CAC's own domestic-language announcements are not consistently archived in English and are accessed here through official state media republication rather than through unofficial translation. Peer-reviewed scholarship includes journal articles on Chinese platform governance, Chinese feminist digital media, Pakistani digital authoritarianism, and Pakistani gender-based digital harm, identified through targeted search on each campaign, statute, and documented event named in the analysis. Non-governmental and international-organization reports include documentation from the Digital Rights Foundation, Amnesty International, the International Federation of Journalists, and Pakistan's National Commission for Human Rights. News reporting, drawn from outlets including Al Jazeera, Dawn, Daily Times, Global Voices, and the South China Morning Post, is used strictly to establish the occurrence, date, and stated official justification of specific events, such as the 2020 TikTok ban or the 2024 blocking of the platform X, rather than to support theoretical or causal claims, which are instead grounded in the peer-reviewed and primary-document literature cited above.

Documents were included where they met three criteria: direct documentation of a named regulatory action, campaign cycle, or legal provision; traceability to an identifiable institutional or authorial source; and, where the document supported a substantive empirical claim rather than merely an event date, corroboration by at least one independent source before being treated as established in the analysis. The corpus spans 2016, the year of PECA's original enactment, through mid-2026, capturing both countries' most recent legal and campaign cycles at the time of writing. In total, the corpus consisted of approximately 43 documents: roughly 13 primary regulatory and documentary texts (statutes, official campaign announcements, and news reports establishing specific regulatory events) and approximately 30 secondary sources (peer-reviewed scholarship and non-governmental or international-organization reports).

## Analytical Procedure

The three analytical axes used in Section 4, definitional, institutional, and operational, were developed inductively by comparing the vocabulary, cited authority, and enforcement pattern documented across the Chinese and Pakistani corpora, and then formalized as a common framework applied uniformly to both cases. The fourth, gendered axis examined in Section 5 emerged from an observation made during this coding process: women-related content and accounts recurred disproportionately across categories that were not, on their face, gender-specific, obscenity and vulgarity in the Pakistani corpus, gender antagonism and fan-circle chaos in the Chinese corpus, prompting a targeted secondary review of that subset of documents. This sequencing, general framework first, gendered pattern identified inductively within it, is intended to reduce the risk that the analysis was structured from the outset to find a gendered convergence that the evidence does not independently support.

## Digital Pardah: Theoretical Originality

The concept of digital pardah developed here is distinguished from six adjacent, and frequently conflated, concepts. Digital surveillance describes a state's or platform's general capacity to observe and record behavior; it does not by itself explain why the object of observation would be gendered, since surveillance capacity is, in principle, applied uniformly across users. Platform governance (Gillespie, 2018) describes the general architecture through which platforms make and enforce rules, treating gender, where addressed at all, as one content category among many rather than as an organizing logic of the system. Visibility politics and algorithmic visibility (Bucher, 2012) explain differential promotion and suppression through engagement-driven ranking, but this literature developed primarily around commercial metrics rather than state-imposed morality categories, and so does not by itself explain why state actors would specifically target women's visibility rather than any other engagement-maximizing content. Algorithmic governance describes the delegation of governing functions to automated systems; digital pardah is compatible with, but not reducible to, this delegation, since much of the enforcement documented in this paper, Pakistani court orders and PTA directives in particular, remains manually adjudicated rather than automated. Digital modesty, where the term appears in existing scholarship, typically describes individual women's self-presentation strategies online; digital pardah instead describes an institutional and infrastructural arrangement imposed on women collectively, independent of their individual self-presentation choices. Gendered moderation and platform patriarchy (Liao, 2024) name the outcome, that moderation disproportionately burdens women, without specifying the joint mechanism through which that outcome is produced. Governmentality (Foucault, 1977) supplies the underlying logic of dispersed, productive power on which digital pardah draws, but Foucault's own account was developed without treating gender as an organizing axis of discipline, a limitation feminist scholars have long identified.

Digital pardah's distinct contribution is to specify that in both cases examined here, no single actor produces the reconstructed modesty order on its own. Removing any one of three actors, state, platform, or family and community, from the account leaves the observed pattern under-explained. State law alone cannot account for informal, at times fatal, violence against visible women in Pakistan, since neither the 2016 killing of Qandeel Baloch nor the 2025 killing of Sana Yousaf followed directly from any statutory provision (Digital Rights Foundation, 2026; Al Jazeera, 2025). Platform policy alone cannot account for why Chinese feminist accounts specifically, rather than all accounts with comparable engagement levels, are targeted under gender-antagonism enforcement (Rest of World, 2023; South China Morning Post, 2025). Family or community authority alone cannot account for the involvement of formal legal categories such as obscenity or gender antagonism in what would otherwise be purely informal social sanction. Digital pardah names the configuration in which these three sources of authority operate jointly, and in the cases examined here largely without formal coordination between them, to reconstruct modesty governance for the digital era. By conceptualizing modesty governance as the emergent product of state regulation, platform infrastructure, and informal social authority acting jointly, digital pardah offers explanatory leverage unavailable in existing theories that privilege only one of these domains.

## Operationalizing Digital Pardah: Analytical Dimensions and Indicators

To move digital pardah from a conceptual claim toward a framework a future researcher could apply to code a new case, three analytical dimensions and associated indicators are proposed here. The first, definitional capture, is indicated by the presence of ostensibly gender-neutral moderation categories, such as vulgarity, decency, or

antagonism, whose documented enforcement nonetheless falls disproportionately on women's accounts or content relative to their overall share of users in the relevant category. The second, joint enforcement, is indicated by evidence that formal, statutory, or platform-level sanction against a woman's visibility co-occurs, in the same case or account, with extra-legal enforcement by family, community, or religious-political actors, without requiring that the two forms of enforcement be coordinated by a single authority. The third, uncoordinated convergence, is indicated by the absence of any documented administrative or legal mechanism linking state, platform, and informal authority, such that their combined effect on a woman's visibility arises from three independently operating logics rather than from a single directed policy. A case exhibiting only one or two of these three dimensions would suggest a related but distinct phenomenon, gendered moderation bias without the family or community dimension, for example, rather than digital *pardah* as specified here; the concept is intended to apply specifically where all three dimensions are jointly present.

### **Strengths and Limitations of Documentary Analysis**

Documentary comparative analysis has particular strengths for this research question. It permits systematic comparison across cases where field access is either constrained, as in Pakistan's contested security environment, or effectively closed to independent researchers, as in China. It draws on a publicly verifiable textual record rather than researcher-elicited testimony, and it allows direct analysis of each state's own justificatory language rather than inference about unstated intent. The method also has clear limitations. It cannot capture how individual women content creators, activists, journalists, or the platform moderators who apply these categories in practice themselves experience, anticipate, or navigate these regimes, a question that would require interview, survey, or digital-ethnographic data beyond this paper's documentary and conceptual scope. The documentary record is shaped by what institutions and journalists chose to document in the first place, introducing a selection effect that cannot be fully corrected for within a documentary design. Finally, a substantial share of Chinese-language primary material is accessed here through English-language state media republication and academic translation rather than through direct engagement with Chinese-language sources, which may filter or soften original terminology in ways not fully recoverable from the secondary record. These limitations mean the claims that follow are best read as conceptual and pattern-level rather than as precise estimates of prevalence.

### **Comparative Analysis**

#### **Defining Harm: Stability, Sanctity, and the Moral Grammar of Regulation**

China's Qinglang campaigns articulate harm primarily through the vocabulary of social stability and collective wellbeing rather than through an explicit sexual or religious morality. Successive campaign cycles have targeted pornography and vulgar livestreaming, cyberbullying and fan-circle conflict, the ostentatious flaunting of wealth, negative emotional content such as pessimism about marriage and childbirth, and, more recently, the misuse of generative artificial intelligence to produce disinformation or bulk low-quality content (Shao & Ao, 2025; ICLG, 2025). This vocabulary is consistent with Feldstein's (2021) broader typology of digital repression, in which social manipulation and content-based control function as instruments of regime stability rather than as claims about individual sexual propriety. Even content concerning gender relations is framed in terms of its capacity to incite antagonism between groups, positioning gender discourse itself as a potential threat to social order rather than as a matter of individual morality.

Pakistan's regulatory language, by contrast, explicitly fuses religious, political, and sexual categories of harm within a single statutory clause. Section 37 of PECA 2016 authorizes removal of content that offends the glory of Islam, threatens the integrity, security, or defence of Pakistan, or violates public order, decency, or morality (Government of Pakistan, 2016). In practice, this clause has been invoked to justify the blocking of blasphemous material, pornography, and politically critical speech under one undifferentiated rubric, reflected in an April 2024 disclosure to the Peshawar High Court that 1.07 million web links had been blocked as blasphemous or pornographic (Dawn, 2024). The 2020 ban of TikTok illustrates this fusion directly: the PTA justified the ban by citing immoral and indecent content, language that made no analytic distinction between religious offense and sexual impropriety (Al Jazeera, 2020).

The two systems are similar in that each frames harm as a threat to a collective order rather than as a private wrong between individuals. They differ in what that order is understood to consist of: a technocratic, largely secular register of social stability in China, and a theocratic, constitutionally entrenched register of religious sanctity in Pakistan. This difference matters sociologically because it determines where contestation over content moderation becomes possible. In China, appeals against moderation can be framed in terms of social benefit or public order, categories that remain, at least nominally, open to negotiation. In Pakistan, appeals must contend with the far less negotiable register of religious sanctity, which is, in Durkheimian terms, sacred rather than merely regulatory (Durkheim, 1893/1984). For the concept of digital pardah, this difference suggests that the statutory language legitimizing modesty governance need not itself be religious for the pattern to appear; a secular vocabulary of stability and harmony is, on the evidence examined here, equally capable of producing it.

### **Who Decides: Bureaucratic Centralization versus Contested Authority**

China's system for adjudicating online harm is highly centralized. The Cyberspace Administration of China operates as the principal national regulator, coordinating campaigns, issuing directives to platforms such as Weibo, Douyin, and WeChat, and reporting large volumes of removed content and suspended accounts in each annual campaign cycle (Wang et al., 2025; Shao & Ao, 2025). Platforms are drawn into this structure as compliance partners rather than independent arbiters, required to implement CAC directives, label AI-generated content, and pre-review user comments before publication (Xinhua, 2025; ICLG, 2025).

Pakistan's institutional landscape is, by contrast, fragmented and increasingly contested. The PTA historically held primary responsibility for blocking content under PECA 2016, but courts have also intervened directly, as when the Peshawar High Court in 2024 ordered the PTA to remove blasphemous material from TikTok and to report back on compliance within seven days (Daily Times, 2024). The 2025 amendment to PECA (Government of Pakistan, 2025) has further fragmented this landscape by creating a new Social Media Protection and Regulatory Authority alongside a Social Media Protection Tribunal and a Social Media Complaint Council, each with overlapping powers to register platforms, investigate complaints, and order content removed (Information Technology and Innovation Foundation, 2025; Khalid Zafar & Associates, 2025). This expansion of regulatory authority has drawn criticism from international rights monitors for its vague definitions of false or offensive content and its limited avenues of judicial appeal (Amnesty International, 2025; International Federation of Journalists, 2025).

Both systems are, on paper, formally centralized: a single national regulator in China, a single originating statute and a small number of named authorities in Pakistan. In practice they diverge sharply. China's centralization is realized operationally through routine bureaucratic compliance, producing a single, traceable chain of authority. Pakistan's nominal centralization is repeatedly supplemented, and at times effectively overridden, by courts, religious-political pressure, and now a proliferating set of overlapping tribunals. This difference matters because it determines whether a targeted account or creator has any predictable venue of appeal, a possibility largely foreclosed in the Chinese case and unevenly, if rarely successfully, available in the Pakistani one. For digital pardah, the comparison suggests that institutional fragmentation does not protect women's digital visibility from regulatory attention; if anything, the multiplication of authorities empowered to act in Pakistan appears to multiply the occasions on which that attention can be triggered.

### **Enforcement Logics: Systemic Algorithmic Governance versus Episodic Reactive Bans**

China's enforcement architecture is increasingly systemic and pre-emptive. Platforms are required to pre-review user comments, label AI-generated content, and maintain moderation infrastructure capable of responding to CAC directives on short notice. The Qinglang campaigns operate on a predictable annual and thematic cycle, a Spring Festival cycle targeting vulgar entertainment and wealth-flaunting, a summer cycle protecting minors, and periodic special actions targeting emergent concerns such as generative AI misuse, such that enforcement resembles scheduled maintenance rather than a series of discrete crises (Shao & Ao, 2025). This systemic quality is reinforced by the requirement that generative AI systems be licensed by the CAC prior to public release, extending pre-emptive control to content that does not yet exist (ICLG, 2025).

Pakistan's enforcement pattern is markedly episodic and reactive. Rather than continuous algorithmic pre-screening, enforcement has historically taken the form of blunt, platform-wide interventions triggered by specific controversies: the multi-year ban of YouTube following a single blasphemous video in 2012, the 2020 ban of TikTok following accumulated public complaint, temporary blanket restrictions on several major platforms in 2021 for unspecified public order reasons, and the 2024 blocking of the platform X under the Removal and Blocking of Unlawful Online Content Rules (Al Jazeera, 2020; TechPolicy.Press, 2024). Even the 2025 PECA amendment, despite creating new regulatory bodies, continues to rely on complaint-driven and directive-based mechanisms rather than any equivalent to China's systemic pre-review infrastructure (Information Technology and Innovation Foundation, 2025).

The two enforcement logics differ in tempo and predictability but converge in one respect relevant to this paper's argument: both generate a recurring category, gender antagonism in China, obscenity and vulgarity in Pakistan, capacious enough to be invoked against a wide range of unanticipated content. In China's systemic model, creators can anticipate the thematic focus of a campaign cycle in advance; in Pakistan's episodic model, they typically cannot. This difference in predictability matters because it changes how governable modesty becomes: in the Chinese case, women content creators can and evidently do adjust output ahead of a known campaign, whereas in the Pakistani case, exposure is discovered only after the fact, when an account or platform is suddenly targeted following an unrelated controversy or shift in political pressure (Amnesty International, 2025).

This difference in tempo also reflects a difference in the underlying moderation apparatus and in the role played by the platform companies themselves. China's pre-review and AI-licensing requirements shift adjudication toward automated classification prior to publication, consistent with Gorwa et al.'s (2020) broader account of algorithmic content moderation as a technical process that encodes political judgments about acceptable speech into classifier design choices that are rarely disclosed or open to contestation by the users they affect. Platforms such as Weibo, Douyin, and WeChat function in this model as compliance infrastructure, building and operating the classifiers CAC directives specify rather than adjudicating content independently. Pakistan's enforcement, by contrast, remains overwhelmingly a matter of human adjudication, court orders, PTA directives, and complaint review, rather than automated flagging, which is consistent with Roberts's (2019) account of content moderation as fundamentally a form of labor performed by reviewers who exercise discretion under conditions the public rarely observes. This distinction matters for digital *pardah* because it locates the site of gendered judgment differently in each case: in China, gendered outcomes are substantially encoded in advance into the classifier categories platforms are required to apply, such as gender antagonism, whereas in Pakistan they remain substantially a matter of case-by-case discretion exercised by court officials, PTA reviewers, and other human decision-makers applying a broadly worded decency standard.

## **GENDERED IMPLICATIONS: WOMEN'S VISIBILITY AND THE DIGITAL PUBLIC SPHERE**

The preceding analysis has shown that China and Pakistan differ fundamentally in regulatory architecture, definitional vocabulary, and institutional authority. The question becomes whether those differences produce equally divergent gendered outcomes. The comparison below suggests they do not: beneath their distinct regulatory grammars, both systems converge on women's digital visibility as a particular object of governance.

Despite substantial differences in definitional grammar, institutional authority, and enforcement logic, both regulatory systems are consistent with a pattern of treating women's digital visibility as a particular object of governance. In China, this is visible in the treatment of feminist speech as inherently destabilizing. The permanent removal of the Feminist Voices Weibo account in 2018, following its publication of International Women's Day content, exemplifies a pattern in which feminist commentary, rather than being distinguished from misogynist harassment, is folded into the same category of socially disruptive antagonism (Rest of World, 2023). This recurred in 2025, when Weibo, acting under the gender-antagonism enforcement priority, closed accounts associated with both misogynist rhetoric and outspoken feminist commentary, treating both as symmetrical threats to social harmony (South China Morning Post, 2025). Wu and Dong (2019) complicate a straightforwardly state-versus-feminists reading of this pattern, showing that Chinese feminism is itself fragmented by class, such that state scrutiny may fall unevenly on a particular, urban and internationally networked, fraction of a more heterogeneous feminist field. Scholarly analysis of the resulting condition describes it as a form of enforced digital disappearance, in which activists persist through repeated account

creation and informal networks rather than sustained public visibility (Shao & He, 2024). The documentary record consulted here describes this persistence largely at the level of pattern, repeated account creation, informal networks, rather than the level of practice, and does not itself establish which specific evasive tactics, such as coded language or migration to smaller, less algorithmically visible spaces, Chinese feminist creators employ; establishing that would require the digital-ethnographic data identified in Section 3.5 as a limitation of the present design.

In Pakistan, the gendering of content moderation operates through a related but distinct mechanism: the conflation of women's bodily visibility with religious and cultural indecency under PECA's decency and morality clauses. The 2020 TikTok ban, framed around obscenity and vulgarity, disproportionately affected a platform whose most visible Pakistani users included young women content creators, several of whom had become targets of moral criticism specifically for their visibility rather than for the content of their speech (Al Jazeera, 2020). This legal framing coexists with, and appears to interact with, extra-legal enforcement of modesty by family and community: the 2016 killing of Qandeel Baloch by her brother, and the 2025 killing of Sana Yousaf, illustrate how the same visibility that draws regulatory scrutiny at the level of the state is also associated with lethal scrutiny at the level of family and community (Digital Rights Foundation, 2026; Al Jazeera, 2025). Naseer and Ashraf (2022) argue, applying the CEDAW framework, that Pakistan's weak internet-security mechanisms independently contribute to this exposure, since inadequate platform-level protection leaves visible women's accounts more vulnerable to coordinated harassment regardless of content. Documentation from the Digital Rights Foundation's helpline is consistent with this account, indicating that women constitute a majority of reported harassment victims, with 58.5 percent of complainants in 2023 and 68 percent of callers in an earlier five-year report, and that women journalists in particular report sexualized harassment intended to damage professional credibility and to pressure self-censorship (Digital Rights Foundation, 2024, 2026). As in the Chinese case, the sources consulted here document harassment and its consequences more fully than the adaptive strategies, pseudonymous branding, migration to private or invitation-only groups, or reliance on digital-rights helplines for rapid-response support, that targeted women may use to manage continued visibility; this asymmetry in the documentary record is itself evidence for the case, made in Section 3.5 and revisited in the Conclusion, for pairing this analysis with interview or digital-ethnographic data.

Read together, these two cases suggest that the gendered structuring of content moderation is not incidental to either system but is instead a structural tendency that persists across sharply different regulatory architectures. In China, the mechanism is primarily definitional: the state's own vocabulary of gender antagonism treats feminist critique and misogynist harassment as equivalent disruptions of harmony, a framing that, per Liao's (2024) analysis of the platformization of misogyny, may in practice burden organized feminist speech more than the harassment it was nominally designed to address. In Pakistan, the mechanism is a blend of legal and extra-legal enforcement, in which decency and morality clauses supply a legal vocabulary that legitimizes, without directly causing, a broader social enforcement of modesty in which family, religious-political actors, and, at times, violence supplement formal regulation. In both cases, women's online presence becomes a site at which the state's claim to guard collective moral order is enacted, producing a reconstituted modesty regime here termed digital *pardah*, not the simple migration of an offline veiling practice into online space, but a new configuration of visibility management co-produced by algorithmic infrastructure, statutory language, and the informal moral authority of family and community. Where a woman's modesty was once negotiated primarily through the physical veil and the immediate social gaze, it is now additionally negotiated through platform policy, state directive, and the possibility of automated or complaint-triggered removal. This shift changes not the fact of moral surveillance but its infrastructure and its scale.

## CONCLUSION

This paper has compared the moral grammar of content moderation in China and Pakistan across three axes, definitional, institutional, and operational, and finds that despite sharp divergence on each, both systems converge in a fourth, cross-cutting respect: disproportionate attention to women's digital visibility. China's Qinglang campaigns define harm through social stability and enforce it through an increasingly systemic, pre-emptive, centralized apparatus; Pakistan's PECA framework defines harm through a fused register of religious sanctity and public decency, enforced through a fragmented and often reactive combination of regulatory bodies, courts, and religious-political pressure. Gender-related speech and women's bodily visibility, in both cases,

attract a disproportionate share of regulatory and, in the Pakistani case, extra-legal attention, suggesting that the governance of women's presence in digital space is not a peripheral effect of content moderation but one of its structural functions.

Theoretically, these findings extend digital pardah beyond a single national or religious context, indicating that the reconstruction of modesty in digital space can occur under a secular-collectivist rationale as readily as under an explicitly religious one. The mechanism identified here, the joint and largely uncoordinated action of state, platform, and informal authority on the same body, is not obviously specific to China or Pakistan. Comparable dynamics have been reported, in preliminary and largely journalistic form, in content moderation debates in Iran, India, and Turkey, where religious-nationalist morality clauses have similarly singled out women's visibility for regulatory attention. Whether digital pardah, as specified in Section 3.4, rather than a looser family resemblance, holds in these additional cases is a question for future comparative work and is not a claim made here.

Methodologically, this paper suggests that documentary comparative analysis, structured through a most-different systems design, is a workable strategy for studying content regulation in contexts where field access is constrained or closed to independent researchers. A two-case design of this kind maximizes analytical leverage over the reach of a convergence, testing whether it survives extreme institutional variation, but does not establish how common the pattern is across the wider population of content-moderating states; extending the comparison to four or more cases with varying regime types and religious establishments would strengthen the inferential warrant considerably. A concrete next step would be to extend the design specifically to Iran, India, and Turkey, coding each against the same definitional, institutional, and operational axes used here, since preliminary and largely journalistic reporting already suggests comparable morality-based targeting of women's digital visibility in each of these settings.

For platform policy and regulation, the analysis indicates that moderation categories framed in gender-neutral language, vulgarity, decency, antagonism, can nonetheless produce gendered enforcement outcomes. Regulators and platforms auditing moderation systems for disparate impact might focus specifically on categories referencing visibility, appearance, or relational conduct, since these correlate more closely with the gendered pattern identified here than categories referencing political or religious content narrowly defined. The paper's limitations should temper this recommendation: the analysis relies on documentary sources rather than primary interview or survey data, a substantial share of Chinese-language material is accessed through English-language secondary reporting, and the two-case design prioritizes analytical reach over population-level generalizability.

A natural extension of this work would pair the conceptual claims made here with an empirical, cross-national study of how women content creators, activists, journalists, and platform moderators themselves experience, anticipate, and negotiate these differing regulatory logics, drawing on semi-structured interviews and digital ethnography with each of these differently positioned groups, a study that would require interview, survey, or digital-ethnographic data beyond this paper's documentary and conceptual scope. Digital pardah, developed through two cases chosen for their institutional dissimilarity, describes a mechanism rather than a fixed national tradition. Wherever platform infrastructure, statutory morality language, and informal community authority act on the same body without formal coordination among them, something resembling digital pardah appears likely to be at work, whether or not the vocabulary of pardah, purdah, or veiling exists locally at all.

## REFERENCES

1. Abu-Lughod, L. (1986). *Veiled sentiments: Honor and poetry in a Bedouin society*. University of California Press.
2. Ahmed, Z. S., et al. (2024). Contestations of internet governance and digital authoritarianism in Pakistan. *International Journal of Politics, Culture, and Society*. <https://doi.org/10.1007/s10767-024-09493-2>
3. Al Jazeera. (2020, October 9). Pakistan bans TikTok over 'indecent' content. Al Jazeera. <https://www.aljazeera.com/news/2020/10/9/pakistan-bans-tiktok-over-indecent-content>
4. Al Jazeera. (2025, June 4). Who was Sana Yousaf, Pakistani TikTok star shot dead by a gunman? Al Jazeera. <https://www.aljazeera.com/news/2025/6/4/who-was-sana-yousaf-pakistani-tiktok-star-shot-dead-by-a-gunman>

5. Amnesty International. (2025, January 29). Pakistan: Authorities pass bill with sweeping controls on social media. Amnesty International. <https://www.amnesty.org/en/latest/news/2025/01/pakistan-authorities-pass-bill-with-sweeping-controls-on-social-media/>
6. Banet-Weiser, S. (2018). *Empowered: Popular feminism and popular misogyny*. Duke University Press.
7. Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, 14(7), 1164–1180. <https://doi.org/10.1177/1461444812440159>
8. Creemers, R. (2017). Cyber China: Upgrading propaganda, public opinion work and social management for the twenty-first century. *Journal of Contemporary China*, 26(103), 85–100.
9. Daily Times. (2024, June 27). PHC orders PTA to remove blasphemous content from TikTok. Daily Times. <https://dailytimes.com.pk/1204152/phc-orders-pta-to-remove-blasphemous-content-from-tiktok/>
10. Dawn. (2024, April 5). PTA blocks 1.07m 'objectionable' web links, PHC informed. Dawn. <https://www.dawn.com/news/1825730/pta-blocks-107m-objectionable-web-links-phc-informed>
11. Digital Rights Foundation. (2024). Cyber harassment helpline report 2023. <https://digitalrightsfoundation.pk/wp-content/uploads/2024/04/DRFs-Cyber-Harassment-Helpline-Report-2023.pdf>
12. Digital Rights Foundation. (2026). Annual report: Digital security helpline (2025). <https://digitalrightsfoundation.pk/wp-content/uploads/2026/05/2025-Helpline-Annual-Report-1.pdf>
13. Durkheim, E. (1984). *The division of labor in society* (W. D. Halls, Trans.). Free Press. (Original work published 1893)
14. Feldstein, S. (2021). *The rise of digital repression: How technology is reshaping power, politics, and resistance*. Oxford University Press.
15. Foucault, M. (1977). *Discipline and punish: The birth of the prison* (A. Sheridan, Trans.). Pantheon Books.
16. Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
17. Global Times. (2025, January). China's cyberspace watchdog launches month-long campaign to promote positive cyber environment ahead of Spring Festival. Global Times. <https://www.globaltimes.cn/page/202501/1327167.shtml>
18. Global Times. (2026, February). China's top internet regulator launches month-long Spring Festival campaign targeting AI-generated harmful content. People's Daily Online. <http://en.people.cn/n3/2026/0213/c90000-20425957.html>
19. Global Voices. (2020, October 23). Pakistan lifts ban on TikTok with conditions. Global Voices. <https://globalvoices.org/2020/10/23/pakistan-lifts-ban-on-tiktok-with-conditions/>
20. Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), 1–15. <https://doi.org/10.1177/2053951719897945>
21. Government of Pakistan. (2016). *Prevention of Electronic Crimes Act, 2016*.
22. Government of Pakistan. (2025). *Prevention of Electronic Crimes (Amendment) Act, 2025*.
23. Han, R. (2016). From vagueness to clarity? Articulating legal criteria of digital content regulation in China. *Global Media and Communication*, 12(3), 211–227.
24. ICLG. (2025). China's key developments in artificial intelligence governance in 2025. In *Telecoms, media and internet laws and regulations 2026. International Comparative Legal Guides*. <https://iclg.com/practice-areas/telecoms-media-and-internet-laws-and-regulations/03-china-s-key-developments-in-artificial-intelligence-governance-in-2025>
25. Information Technology and Innovation Foundation. (2025, May 16). Pakistan's content moderation regulation. ITIF. <https://itif.org/publications/2025/05/16/pakistan-content-moderation-regulation/>
26. International Federation of Journalists. (2025, January 29). Pakistan: PECA amendments further tighten government grip on digital expression. IFJ. <https://www.ifj.org/media-centre/news/detail/article/pakistan-peca-amendments-further-tighten-government-grip-on-digital-expression>
27. Jamil, S. (2021). The rise of digital authoritarianism: Evolving threats to media and Internet freedoms in Pakistan. *World of Media: Journal of Russian Media and Journalism Studies*, 3, 5–33. <https://doi.org/10.30547/worldofmedia.3.2021.1>

28. Khalid Zafar & Associates. (2025). Prevention of Electronic Crimes (Amendment) Act, 2025. <https://khalidzafar.com/laws-of-pakistan/prevention-of-electronic-crimes-amendment-act-2025/>
29. Liao, S. (2024). The platformization of misogyny: Popular media, gender politics, and misogyny in China's state-market nexus. *Media, Culture & Society*, 46(1), 191–203. <https://doi.org/10.1177/01634437221146905>
30. MacKinnon, R. (2011). China's "networked authoritarianism." *Journal of Democracy*, 22(2), 32–46. <https://doi.org/10.1353/jod.2011.0033>
31. Mahmood, S. (2005). *Politics of piety: The Islamic revival and the feminist subject*. Princeton University Press.
32. Naseer, S., & Ashraf, C. (2022). Gender-based violence in Pakistan's digital spaces. *Feminist Legal Studies*, 30(1), 29–50. <https://doi.org/10.1007/s10691-021-09473-3>
33. National Commission for Human Rights (Pakistan). (2026). Report on PECA and the 2025 amendments act. <https://nchr.gov.pk/wp-content/uploads/2026/02/NCHR-Report-on-PECA-and-the-2025-Amendments-Act.pdf>
34. Przeworski, A., & Teune, H. (1970). *The logic of comparative social inquiry*. Wiley-Interscience.
35. Rest of World. (2023). China's internet feminist movement thrives despite censorship. <https://restofworld.org/2023/china-online-feminist-movement/>
36. Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press.
37. Roberts, T., & Oosterom, M. (2024). Digital authoritarianism: A systematic literature review. *Information Technology for Development*, 1–25. <https://doi.org/10.1080/02681102.2024.2425352>
38. Shao, L., & Ao, M. (2025). Thematic evolution of China's media governance policies: A tri-logic synergistic perspective. *Information*, 16(8), Article 696. <https://doi.org/10.3390/info16080696>
39. Shao, S., & He, G. (2024). After account bombing: Chinese digital feminists haunt platform censorship as cyber living ghosts. *Feminist Media Studies*, 24(5), 944–961. <https://doi.org/10.1080/14680777.2024.2311362>
40. South China Morning Post. (2025). China's Weibo shuts down social media accounts for 'inciting gender antagonism.' <https://www.scmp.com/news/china/politics/article/3322407/chinas-weibo-shuts-down-social-media-accounts-inciting-gender-antagonism>
41. TechPolicy.Press. (2024). Pakistan on verge of techno-authoritarian turn. <https://www.techpolicy.press/pakistan-on-verge-of-techno-authoritarian-turn/>
42. Wang, Y., Md Syed, M. A., & Shamshudeen, R. I. (2025). Cultural governance and censorship: Unpacking China's Qinglang Action. *Asian Journal of Communication*, 35(5), 469–489. <https://doi.org/10.1080/01292986.2025.2526475>
43. Wu, A. X., & Dong, Y. (2019). What is made-in-China feminism(s)? Gender discontent and class friction in post-socialist China. *Critical Asian Studies*, 51(4), 471–492. <https://doi.org/10.1080/14672715.2019.1656538>
44. Xinhua. (2025, February 21). China targets AI misuse in 2025 cyberspace cleanup initiative. Xinhua News Agency. <https://english.news.cn/20250221/c30142348f62494d9cee5d8098061ee5/c.html>