



Selected Papers of #AoIR2024:  
The 25th Annual Conference of the  
Association of Internet Researchers  
Sheffield, UK / 30 Oct - 2 Nov 2024

## **BORDERLINE CONTENT AND PLATFORMISED SPEECH GOVERNANCE: MAPPING TIKTOK'S MODERATION CONTROVERSIES IN SOUTH AND SOUTHEAST ASIA**

Diyi Liu  
University of Oxford

### **Introduction**

Digital platforms hosting user-generated content (UGC) have increasingly become the “prism” that shapes what we see and how we see the world (Flyverbom, 2019). They are algorithmically and institutionally managed assemblages that enable (or constrain) interactions, control visibility, and curate public discourse (Gillespie, 2010, 2018; Helberger et al., 2018). Through technical and normative design, online platforms not only offer sets of affordances and constraints for public discourse, but also, in Weber’s account, seek to establish and enforce maxims or rules toward which social behaviour is orientated; in other words, platforms actively engage in the legitimation of online content governance (Weber, 1964; Zelditch, 2001).

Numerous literature has documented the increasing reliance of platforms’ facilitated algorithmic moderation systems to govern online speech (Balkin, 2018; Caplan, 2023; Gillespie, 2018; Gorwa, 2019; Grimmelmann, 2015). *Platformised speech governance*, as termed in this paper, refers to the governance mechanisms by which online speech is regulated and moderated specifically *through* and *by* digital platforms within a particular social context. As part of their increasing role in the public sphere (Fischer & Jarren, 2023), content moderation of UGC platforms involves an ongoing process of legitimation through interactions among platforms and other social actors, who continually rearticulate moderation norms sensitive to each national context. It sheds light on contemporary struggles over the boundaries of free expression online. At the core of the governance mechanisms are questions regarding how much liability these platforms should face, the specific content categories subject to regulation, by what legal basis and moral standards these decisions should be made, and by whom. Often this transcends mere questions of legality, moving into a broader discussion about the

Suggested Citation (APA): Liu D. (2024, October). *Borderline content and platformised speech governance: Mapping TikTok’s moderation controversies in South and Southeast Asia*. Paper presented at AoIR2024: The 25th Annual Conference of the Association of Internet Researchers. Sheffield, UK: AoIR. Retrieved from <http://spir.aoir.org>.

legitimacy of online content governance (Haggart & Keller, 2021). The normative rules set by platforms have become integral to the wider regulatory landscape of online speech. Nevertheless, the legitimacy of these governance efforts is a subject of challenge and varies across different contexts (Suzor et al., 2018).

While extensive research has explored content moderation of digital platforms in democratic contexts, there remains a gap in understanding whether the Western communication order's legitimation tools adequately address the complexities of less-than-democratic developing nations (Ramesh et al., 2022). One significant challenge is that governance discussions often revolve around high-level international human rights frameworks (Jørgensen, 2017), yet do not fully account for the reality that free speech can be and have been interpreted in different ways across the globe, even within democratic societies such as those in Europe and the United States (Riedl et al., 2021). On that front, this paper expands the existing discussion of platformised speech governance through a case study of how TikTok, as a user-generated content platform originating from China, engages in local content moderation in areas that have been overlooked in scholarship - South and Southeast Asia. The case selection provides an analytically broad spectrum for understanding different power dynamics that may be said to shape and influence the legitimacy of transnational platform companies in Asia as they become manifest in very different but regionally related contexts.

## **Research Methods**

TikTok was firstly introduced into the Indonesian market in late 2017, marking the beginning of its presence in the SA/SEA region. The data collected for this study spans from September 2017 to September 2023, though specific data availability varies across sources as indicated in the following section. The primary objective is to map how the platform and local governments draw the lines of content moderation, and how the boundaries of what constitutes "borderline content" have been contested during controversial incidents. To achieve this, this paper pursues an iterative empirical approach while triangulating three key data sources:

*TikTok's moderation rules and enforcement in SA/SEA.* Above all, I examine TikTok's moderation rules and public disclosure. These includes its Terms of Service, Community Guidelines, Branded Content Policy, in particular with jurisdictional specifications in SA/SEA countries. Additionally, I investigate some key metrics in its marketing and transparency data, which partly indicates the platform's rule enforcement, priorities, and compliance. It is acknowledged that the public-facing transparency data has limitations in terms of the actual auditability (Ananny & Crawford, 2018), yet it still provides insights into the temporality and discursive performativity of TikTok's content governance (Chan et al., 2023; Scharlach et al., 2023).

*Legal regulations and government policies.* Following initial examination of TikTok's localised moderation in this region, five key countries are selected for further analysis based on their market potential and moderation landscape: Indonesia, Pakistan, Vietnam, the Philippines, and Thailand. The analysis focuses on legal regulations and government policies specifically in regard to (a) online speech restriction, and (b) platform liability. The aim is to understand the legal and political contexts, if (and how)

TikTok establishes and modulates its rules across the aforementioned jurisdictions, and how local regulatory traditions are encoded into the normative framework of the platform.

*Controversy mapping.* Unlike Google who discloses government requests from judiciary and executive branches, these details cannot be easily extracted in TikTok's current transparency reporting structure. Given that normative documents don't offer detailed and granular moderation policies (e.g., types of content) in different countries, this paper posits that power dynamics come to the forefront through controversy – notable incidents that challenges a platform's core principles, prompting operational adaptations (Ananny & Gillespie, 2017; Marres, 2015). I further explored major reported content moderation controversies in the SA/SEA region. Specifically, I query media reports in Factiva News Database spanning from September 2017 to September 2023. The search was refined to include media outlets from South Asia and Southeast Asia, yielding results including relevant international news outlets as well as local outlets in the five countries. Despite potential limitations stemming from language translation and media coverage bias, high-profile cases or news events were prioritised for analysis due to their tendency to highlight tensions and controversies for initial mapping. Coding was conducted sequentially for each identified controversy case. The following key dimensions were systematically extracted and analysed across cases: (a) time of controversy, (b) controversial content areas, (c) stakeholders involved, (d) regulating approach, and (e) TikTok's responses (if recorded).

## **Findings**

Overall, significant uncertainties exist in setting boundaries for online speech by both TikTok and local governments in SA/SEA. While the platform demonstrates a degree of sensitivity to local contexts, its practices primarily serve its for-profit nature rather than moral obligations. Furthermore, the results highlight TikTok's depoliticisation of content governance, that the platform positions itself as less involved in contentious political debates, so as to justify its inconsistent and highly pragmatic moderation approaches. Meanwhile, states in the region, despite aspiring to maintain a central role in speech governance, struggle with limited technical capabilities in moderating at scale. Their interests in appropriating platformised moderation for political needs, however, often leads to reliance on vague rationales such as securitisation and public morality, as suggested in recurring controversies and decisive actions in regulatory moves. At the heart of these issues lie borderline content areas constantly facing ethical dilemmas and rhetoric misuse. Subjective definitions based on intent and different levels of judicial oversight across countries further contribute to ambiguity in the normative debate. The contradictory goals of (de)politicising borderline moderation seemingly counterbalance each other, yet in practice lead to an accountability vacuum without legitimate interests. Given the lack of normative common ground, the study highlights the significance of procedural justice and civic participation to mitigate rhetoric that rationalises imposition of speech norms hinging on imbalanced political power.

## **References**

- Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989. <https://doi.org/10.1177/1461444816676645>
- Ananny, M., & Gillespie, T. (2017). *Public Platforms: Beyond the Cycle of Shocks and Exceptions*. Oxford Internet Institute. <http://blogs.oii.ox.ac.uk/ipp-conference/sites/ipp/files/documents/anannyGillespie-publicPlatforms-oii-submittedSept8.pdf>
- Balkin, J. M. (2018). Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation. *U.C. Davis Law Review*, 51(3), 1149–1210.
- Caplan, R. (2023). Networked Platform Governance: The Construction of the Democratic Platform. *International Journal of Communication*, 17, 3451–3472.
- Chan, N. K., Su, C. C., & Shore, A. (2023). Shifting platform values in community guidelines: Examining the evolution of TikTok’s governance frameworks. *New Media & Society*, 14614448231189476. <https://doi.org/10.1177/14614448231189476>
- Fischer, R., & Jarren, O. (2023). The platformization of the public sphere and its challenge to democracy. *Philosophy & Social Criticism*, 01914537231203535. <https://doi.org/10.1177/01914537231203535>
- Flyverbom, M. (2019). *The Digital Prism: Transparency and Managed Visibilities in a Datafied World* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781316442692>
- Gillespie, T. (2010). The politics of ‘platforms’. *New Media & Society*, 12(3), 347–364. <https://doi.org/10.1177/1461444809342738>
- Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
- Gorwa, R. (2019). What is platform governance? *Information, Communication & Society*, 22(6), 854–871. <https://doi.org/10/gfwbn4>
- Grimmelmann, J. (2015). The Virtues of Moderation. *Yale Journal of Law and Technology*, 17, 42–109. <https://doi.org/10.31228/osf.io/qwxf5>
- Haggart, B., & Keller, C. I. (2021). Democratic legitimacy in global platform governance. *Telecommunications Policy*, 45(6), 102152. <https://doi.org/10.1016/j.telpol.2021.102152>
- Helberger, N., Pierson, J., & Poell, T. (2018). Governing online platforms: From contested to cooperative responsibility. *The Information Society*, 34(1), 1–14. <https://doi.org/10.1080/01972243.2017.1391913>
- Jørgensen, R. F. (2017). What Platforms Mean When They Talk About Human Rights. *Policy & Internet*, 9(3), 280–296. <https://doi.org/10.1002/poi3.152>
- Marres, N. (2015). Why Map Issues? On Controversy Analysis as a Digital Method. *Science, Technology, & Human Values*, 40(5), 655–686. <https://doi.org/10.1177/0162243915574602>
- Ramesh, D., Kameswaran, V., Wang, D., & Sambasivan, N. (2022). How Platform-User Power Relations Shape Algorithmic Accountability: A Case Study of Instant Loan

Platforms and Financially Stressed Users in India. *2022 ACM Conference on Fairness, Accountability, and Transparency*, 1917–1928.  
<https://doi.org/10.1145/3531146.3533237>

Riedl, M. J., Whipple, K. N., & Wallace, R. (2021). Antecedents of support for social media content moderation and platform regulation: The role of presumed effects on self and others. *Information, Communication & Society*, 1–18.  
<https://doi.org/10.1080/1369118X.2021.1874040>

Scharlach, R., Hallinan, B., & Shifman, L. (2023). Governing principles: Articulating values in social media platform policies. *New Media & Society*, 146144482311565.  
<https://doi.org/10.1177/14614448231156580>

Suzor, N. P., Van Geelen, T., & Myers West, S. (2018). Evaluating the legitimacy of platform governance: A review of research and a shared research agenda. *International Communication Gazette*, 80(4), 385–400. <https://doi.org/10.1177/1748048518757142>

Weber, M. (1964). *The Theory of Social and Economic Organization* (T. Parsons, Ed.; A. M. Henderson, Trans.). Free Press.

Zelditch, M. (2001). Processes of Legitimation: Recent Developments and New Directions. *Social Psychology Quarterly*, 64(1), 4–17. <https://doi.org/10.2307/3090147>