

Content Moderation

Key Considerations

Though the challenges associated with content moderation in the digital environment predate the ascendancy of the cloud, the cloud's scale and spread intensify the existing debate, especially in terms of questions of responsibility and liability for moderation, and introduce some new governance challenges.¹

- **Divergent rules across jurisdictions.** Different jurisdictions have historically opted for different trade-offs in governing online content and setting content moderation standards, resulting in a fragmented environment and driving up the compliance burden.
- **Division of responsibilities among stakeholders.** The division of responsibilities and liabilities between cloud providers and their enterprise customers for the moderation of cloud-hosted content remains somewhat unclear.² While enterprise customers are generally understood to be primarily responsible for moderating the content hosted on their platforms, regulators and the public are increasingly turning to cloud providers to intervene when enterprise customers are unwilling to do so. However, providers generally oppose being held liable for the misuse of their platforms and seek to pass operational and moderation burdens onto platform users and enterprise customers. Governments and customers seek clarity on the division of responsibilities in moderating content and may desire clear and transparent rules to govern cloud providers' moderation policies.
- **Mistargeted takedown requests.** Requests by government regulators or policymakers for cloud providers to remove content may be misdirected due to a lack of understanding of the technical dimensions of cloud infrastructure. For example, cloud providers may not be able to take down specific pieces of their customers' content. As a result, governments might need to consider directing their takedown requests to the owners of the data.
- **Lack of transparency in moderation practices and processes.** Providers are often not required to disclose content moderation practices nor how they approach takedown requests. Moreover, merely auditing content moderation practices might be insufficient without considering the context of other regulatory mechanisms or established industry standards.
- **Few guardrails for developing and using automating content moderation tools.** Requirements for proactive and/or swift moderation of content incentivize the adoption of automated, provider-developed content moderation tools by cloud customers.³ This may pose concerns for the transparency of moderation policies, given that these tools' design, training, and deployment is often opaque. It may also raise concerns about the accuracy of

such tools and given the potential for false positives and negatives associated with automation.

- **Effects on free speech and user behavior.** Recognizing that a small number of cloud providers are responsible for hosting or storing much of the content on the internet, the content moderation policies set by these entities can directly shape the parameters of acceptable speech and user behavior online. This would render cloud infrastructure service providers as internet gatekeepers.⁴
- **Effects on competition.** The precedent being set by infrastructure providers in moderating content increases the already-significant barriers (primarily financial and personnel costs) presented to entrants to the highly concentrated cloud infrastructure market, as it introduces expectations (which may potentially be codified) that they also perform content moderation in addition to the provision of other cloud-related services.
- **Effects on the internet and its architecture.** If cloud infrastructure providers regularly make decisions on what content should and should not be hosted, delivered, stored, and so on, this will in effect contribute to the fragmentation of the internet, potentially impairing its ability to support commerce and communication around the world.
- **Moderation can be controversial.** Upholding political neutrality in policies and practices might be widely preferred by companies that seek to appeal to the widest possible market, yet its implementation is bound to prove controversial and burdensome. Cloud providers might be subject to allegations of bias due to the particulars of how they exercise “neutrality” in content moderation activities (including taking down and leaving up certain content).
- **Dependence of foreign providers.** Concerns about bias in service availability, reliability, and integrity may be further exacerbated in cases of dependence on foreign providers. These concerns may arise from perceptions that content moderation policies and procedures are largely the product of Western-centric perspectives and might not translate to the customs or values of non-Western regions (for example, what constitutes hate speech or inappropriate content for children). They can also arise from a recognition that cloud providers might not offer moderation resources in certain languages.
- **Setting dangerous precedent.** Excessive or inappropriate government requirements that cloud providers moderate content could create precedent for and legitimize authoritarian regimes’ efforts to moderate content in ways that are inconsistent with democratic values.

The capacity of cloud providers to arbitrate political expression and discourse was thrown into sharp relief following Amazon Web Service’s (AWS) suspension of Parler—an online social networking platform used in large part by far-right groups—in January 2021.⁵ The AWS-Parler case raised several concerns regarding perceived political neutrality in content moderation policies and practices, especially if those mandates are influenced by governments or advocacy groups. Issues regarding content and service moderation extend beyond free speech, however, and include cloud governance challenges pertaining to customer data protection, enterprise autonomy, and so on. For example, the abuse of content moderation practices by cloud providers and their enterprise customers might facilitate service bias or worsen unequal access to cloud-based services. Moreover, leadership or regime change might produce onerous shifts vis-à-vis censorship or government intrusion.

These concerns compound the risks of bias, opacity, and inaccuracy introduced by black-box content moderation algorithms. Although the deployment of such technologies and practices might appear benevolent—such as Apple’s recently announced plan to scan iCloud data for child sexual abuse material⁶—the consequences of misidentification can be devastating. In the case of Apple, local authorities are notified once content is flagged by internal systems and reviewers, which might lead to misinformed arrests if verification protocols are flawed or inaccurate. The potential risks to user privacy through increased corporate and state surveillance are self-evident.

Key cloud stakeholders might adopt diverse arrangements in response to these challenges. For example, governments might mandate oversight and transparency on content moderation policies and requests whereas providers might reject unlawful content moderation requests. These arrangements vary in scope, intensity, and temporality and might therefore produce tension between stakeholders. However, despite the vast landscape of challenges presented by content moderation vis-à-vis cloud governance, there remains space for cooperation and advancement across sectors and industries.

Recent Examples

- [“Amazon Web Services disables ISIS propaganda website it had hosted since April,”](#) *The Washington Post*, August 27, 2021
- [“Seldom-seen Amazon unit made the call that brought down pro-Trump Parler,”](#) *The Washington Post*, January 13, 2021
- [“Why Banning 8chan Was So Hard for Cloudflare: ‘No One Should Have That Power’,”](#) *The New York Times*, August 5, 2019
- [“Why We Terminated Daily Stormer,”](#) Cloudflare, August 16, 2017


Notes

¹ There is an ongoing debate about whether the entities that engage in content moderation are “publishers” of the content (and therefore responsible for moderating the content and ameliorating its negative effects) or are conduits of the content (and therefore not responsible for its effects). The status of cloud providers in this taxonomy is currently somewhat unclear.

² For more on this issue, see our analysis of the [Division of Responsibility Among Stakeholders](#).

³ For example, see: “Amazon Rekognition,” Amazon Web Services, n.d., <https://aws.amazon.com/rekognition/?nc=sn&loc=0> and “Content Moderator,” Microsoft Azure, n.d., <https://azure.microsoft.com/en-us/services/cognitive-services/content-moderator/>

⁴ For example, see: Sheila Dang, “EXCLUSIVE Amazon considers more proactive approach to determining what belongs on its cloud service,” Reuters, September 5, 2021,



<https://www.reuters.com/technology/exclusive-amazon-proactively-remove-more-content-that-violates-rules-cloud-2021-09-02/>

⁵ Jay Greene, Rachel Lerman, and Tony Romm, “Seldom-seen Amazon unit made the call that brought down pro-Trump Parler,” *The Washington Post*, January 13, 2021, <https://www.washingtonpost.com/technology/2021/01/13/amazon-parler-takedown/>

⁶ Sarah Morrison, “The controversy over Apple’s plan to protect kids by scanning your iPhone,” *Vox*, August 13, 2021, <https://www.vox.com/recode/2021/8/10/22617196/apple-ios15-photo-messages-scanned>