



## Cloud Access Restrictions and Content Moderation

As cloud providers increasingly facilitate widespread access to online services,<sup>1</sup> they have acquired significant power to directly and indirectly shape the content that appears online, which can have the effect of denying customers access to large amounts of information, the ability to express themselves online,<sup>2</sup> as well as the ability to access essential cloud-hosted applications and public services.<sup>3</sup> They do this directly by, for example, refusing to host certain customer and user content.<sup>4</sup> A cloud provider's decision not to host particular content might be the result of a government request to take down that content or to stop providing service to a particular entity or population; it also could be the result of the provider's own terms of service or general business interests. Cloud providers also indirectly shape the virtual public square by offering targeted content moderation tools,<sup>5</sup> which their enterprise customers may use to identify and remove content on their own services.

Content-moderation issues, of course, precede the ascendance of the cloud. For example, social media companies have long struggled to moderate content on their platforms and constantly grapple with how their decisions affect different communities around the world. Similarly, some governments have long restricted online access through internet shutdowns and prohibitions on accessing particular content and services. While the underlying problem is not new, as cloud providers play a larger role in underpinning critical tools that support global commerce and communications, providers are also playing a greater role in directly and indirectly moderating content and facilitating access to it, and thus have become more central to these debates.

### Key Considerations

- **Absence of public debate, informed agreement, and norms.** There is no robust public debate and consensus to inform norms on acceptable policies for cloud-level content moderation and restriction of access. Moreover, differences of opinion between and within national governments make it challenging to codify a single set of standards and best practices into law.
- **Limited transparency into cloud providers' content moderation and access restriction policies and practices.** The public has little insight into how cloud providers make decisions on content moderation and restricting access, including at the behest of domestic and foreign governments. Some governments may coerce cloud service and application providers to remove targeted content,<sup>6</sup> limit certain populations' access to cloud services, and eschew demands to establish public criteria for doing so.

Lack of clarity regarding content moderation tools and their use. The design, training, and deployment of provider-developed content moderation tools are often opaque.<sup>7</sup>

- **Setting dangerous precedent.** Excessive government requirements that cloud providers restrict access or moderate content could create precedent for and legitimize authoritarian regimes' efforts to selectively restrict access and moderate content in ways that are inconsistent with democratic values.

## Stakeholder Perspectives

### Government

- Seek to maintain jurisdiction over cloud-hosted services and data to ensure that providers and their customers comply with domestic laws.
- Vary internationally and internally in their desires as well as rationales for restricting access and moderating content.
- Vary in their desire to attract investment from foreign cloud providers in the domestic market.<sup>8</sup> Many, however, wish for physical cloud infrastructure to be constructed and operated in their jurisdiction.<sup>9</sup>
- Vary in their desire to ensure that the national population has fair and

### Providers

- Comply with legitimate government requests to restrict access to services and moderate content in a manner that allows them to continue doing business in host countries.
- Seek to maintain customer confidence that their data is not being mismanaged or moderated in a biased fashion.
- Want governments to narrow their rationales for requiring cloud providers to restrict access to cloud services or moderate cloud-hosted content.
- Seek clarity on how their responsibilities differ from those of their customers in

### Customers

- Wish to access cloud services free from discrimination.
- Seek to safeguard (and ensure the providers protect) the privacy and security of their activities and data.
- Want protection from infringements on freedoms, such as political speech, access to information, or economic inclusion. (Similar to human and civil rights advocacy groups' perspective.)
- Desire insight into how cloud providers handle government service restriction and/or takedown requests.
- *Enterprise customers:* Seek clarity on how their responsibilities differ from those of

### Others

- *Human and civil rights advocacy groups:* Seek to ensure that cloud access and usage remains consistent with fundamental rights. (Similar to customers' perspective.)

equitable access to cloud services.

moderating content and restricting access. (Shared with enterprise customers' perspective.)

their cloud providers in moderating content and restricting access. (Shared with cloud providers' perspective.)

- Some want governments to establish clear and transparent rules to govern cloud providers' content moderation and access restriction policies.

## Tensions with Other Cloud Governance Issues

- **Equitable Cloud Access:** Government requirements that cloud providers restrict, deny, or suspend certain populations' access to cloud services and content can adversely impact the ability of these communities to use essential services and express themselves online.

## Potential Ways Ahead

### Government

- Clearly define criteria, standards, and procedures for requiring cloud providers to take down content or deny/suspend service(s).
- Establish mechanisms to audit for bias the content moderation

### Providers

- Leverage best practices on transparency and accountability in restricting access and the development of content moderation tools<sup>11</sup> (for example, the Santa Clara Principles and Digital Services Act).
- Refrain from restricting access or

### Customers

- Familiarize themselves with their data rights and protections against bias.
- Notify relevant public entities when suspected instances of bias in content moderation do occur.

### Others

- *Independent, third-party corporate auditors:* Carry out algorithmic audits of cloud providers' and their customers' content moderation technologies.

- technologies being deployed or offered by cloud providers.
- Establish clear and easy to use pathways for customers seeking redress following suspected instances of bias.
  - Encourage periodic publication of algorithmic impact statements and content moderation.
  - Incentivize major cloud providers to solicit public feedback on the development and implementation of content moderation policies, practices, and technologies.
  - Develop standards for oversight, accountability, and transparency of providers' policies and content moderation requests.
  - Model domestic rules regarding service availability after principles such as Article 19 of the Universal
- moderating content in ways that do not meet agreed-upon public standards and procedures.
- Implement "notice-and-comment" procedure before adopting/updating content moderation policies, practices, and technologies.
  - Provide transparency into content moderation policies, as well as requests<sup>12</sup> and how providers comply with them, in every circumstance permitted by law (that is, via permanent/temporary removal of content/suspension of services).
  - Establish a review process whereby customers and end-users can challenge content removal by providers, the ability to do so is clear, and the procedure thereof is straightforward.
  - Agree to adopt and promote the use of explainable algorithms<sup>13</sup> in their products as well as to subject their content moderation offerings to algorithmic audits by independent, third-party corporate
- *Consumer customers:* Use redress mechanisms made available by providers and government.
  - *Enterprise customers:* Solicit public feedback before adopting/updating content moderation policies, practices, and technologies.
  - *Enterprise customers:* Create internal review boards responsible for reviewing content moderation decisions.<sup>17</sup>
  - *Enterprise customers:* Ensure affected parties can challenge content removal, that the ability to do so is clear, and the procedure thereof is straightforward.
  - Where possible and effective, use bargaining power (for example, threatening to move or stop using platforms or providers) to compel providers to be more

Declaration of Human Rights.<sup>10</sup>

- auditors.<sup>14</sup> (Shared with independent, third-party corporate auditors.)
  - Publish regular algorithmic impact statements.<sup>15</sup>
  - Ensure developers of content moderation tools are trained on machine bias, diversity, inclusion, and equity issues.<sup>16</sup>
- transparent and accountable.
  - Establish customer networks to facilitate discourse and disseminate information regarding changes in governmental and/or corporate policy and practices.

## Recent Examples

- [“Amazon suspends Parler, taking pro-Trump site offline indefinitely,”](#) *The Washington Post*, January 11, 2021.

## Notes

<sup>1</sup> While this project recognizes that major social media platforms are functionally SaaS providers, this section does not treat them as “cloud providers” in order to focus on the unique ways cloud computing technologies fit into the broader debate over content moderation.

<sup>2</sup> United Nations, “Universal Declaration of Human Rights,” United Nations, December 10, 1948, <https://www.un.org/en/about-us/universal-declaration-of-human-rights#:~:text=Article%2019,media%20and%20regardless%20of%20frontiers>.

<sup>3</sup> Ibid.

<sup>4</sup> This includes de-platforming, as in the case of AWS and Parler (Tony Romm and Rachel Lerman, “Amazon suspends Parler, taking pro-Trump site offline indefinitely,” *The Washington Post*, January 11, 2021, <https://www.washingtonpost.com/technology/2021/01/09/amazon-parler-suspension/>), as well as takedowns, labeling and other traditional forms of content moderation as practiced by social media companies.

<sup>5</sup> Amazon Web Services, “Amazon Rekognition,” Amazon Web Services, n.d., <https://aws.amazon.com/rekognition/?blog-cards.sort-by=item.additionalFields.createdDate&blog-cards.sort-order=desc> and Microsoft Azure, “Content Moderator,” Microsoft Azure, n.d., <https://azure.microsoft.com/en-us/services/cognitive-services/content-moderator/>.

<sup>6</sup> For example, see: Campbell Kwan, “Twitter labels India’s new content blocking powers as threat to freedom of expression,” ZDNet, May 27, 2021, <https://www.zdnet.com/article/twitter-labels-indias-new-content-blocking-powers-as-threat-to-freedom-of-expression/>.

<sup>7</sup> Amazon Web Services, “Amazon Rekognition,” Amazon Web Services, n.d., <https://aws.amazon.com/rekognition/?blog-cards.sort-by=item.additionalFields.createdDate&blog-cards.sort-order=desc> and Microsoft Azure, “Content Moderator,” Microsoft Azure, n.d., <https://azure.microsoft.com/en-us/services/cognitive-services/content-moderator/>.

<sup>8</sup> Qatar Financial Centre, “Qatar Remains Open for Business,” Bloomberg, <https://sponsored.bloomberg.com/immersive/qatar-financial-centre/qatar-open-business>.


<sup>9</sup> Amazon Web Services and Intel, “Meeting your Data Residency Requirements,” Amazon Web Services and Intel, n.d., <https://d1.awsstatic.com/product-marketing/Outposts/AWS%20Data%20Residency%20Infographic.pdf>.

<sup>10</sup> Catherine Howell and Darrell M. West, “The internet as a human right,” *The Brookings Institution*, November 7, 2016, <https://www.brookings.edu/blog/techtank/2016/11/07/the-internet-as-a-human-right/>.

<sup>11</sup> For example, the Santa Clara Principles lay out baseline standards on transparency, notice, and appeal, that companies engages in content moderation may subscribe to (Content Moderation and Removal at Scale, “The Santa Clara Principles on Transparency and Accountability in Content Moderation,” Content Moderation and Removal at Scale, May 7, 2018, <https://santaclaraprinciples.org/>). Likewise, Article 23 of the EU’s Digital Services Act (European Commission, “The Digital Services Act: ensuring a safe and accountable online environment,” European Commission, December 15, 2020, [https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment\\_en](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en)), which calls for transparency on the use of automated moderation tools, including transparency on what the precise purpose of the tool is as well as indicators of the accuracy of its filters and safeguards against error.

<sup>12</sup> Though not a cloud service provider, Twitter offers a useful framework for delivering insight into content moderation requests levied by governments: Twitter, “Removal Requests,” Twitter Transparency, n.d., <https://transparency.twitter.com/en/reports/removal-requests.html#2020-jul-dec>.

<sup>13</sup> The auditability of AI remains contentious due to the black-box nature of these systems as well as to the security and commercial concerns by providers over auditing the source code for their technologies. A potential avenue to consider may be the adoption of explainable artificial intelligence (XAI) algorithms, which follow the three principles of transparency, interpretability, and explainability. In doing so, auditors and end-users may be better able to examine the



systems and determine how it is making decisions and whether the results of these decisions are as expected. See: Amina Adadi and Mohammed Berrada, “Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI),” in *IEEE Access* 6, (Fall 2018): 52138—60, <https://ieeexplore.ieee.org/document/8466590>.

<sup>14</sup> While such audits are relatively uncommon given the sensitivity around providers’ proprietary software, they are growing in popularity. Alfred Ng, “Can Auditing Eliminate Bias from Algorithms,” *The Markup*, February 23, 2021, <https://themarkup.org/ask-the-markup/2021/02/23/can-auditing-eliminate-bias-from-algorithms> and Rumman Chowdhury and Jutta Williams, “Introducing Twitter’s first algorithmic bias bounty challenge,” *Twitter Engineering* (blog), July 20, 2021, [https://blog.twitter.com/engineering/en\\_us/topics/insights/2021/algorithmic-bias-bounty-challenge](https://blog.twitter.com/engineering/en_us/topics/insights/2021/algorithmic-bias-bounty-challenge).

<sup>15</sup> Nicol Turner Lee, Paul Resnick, and Genie Barton, “Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms,” *The Brookings Institution*, May 22, 2019, <https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>.

<sup>16</sup> For example, see how developers can better understand and practice racial sensitivity: Jessie Daniels, Mutale Nkonde, Darakhshan Mir, *Advancing Racial Literacy in Tech* (New York City: Data & Society, 2019), [https://datasociety.net/wp-content/uploads/2019/05/Racial\\_Literacy\\_Tech\\_Final\\_0522.pdf](https://datasociety.net/wp-content/uploads/2019/05/Racial_Literacy_Tech_Final_0522.pdf).

<sup>17</sup> See, for example, Facebook’s Oversight Board: Oversight Board, “Oversight Board Home Page,” Oversight Board, n.d., <https://oversightboard.com/>.