

Aapti Institute, June 2026 | **All Things in Moderation 2026**

Reflecting on our session and on principles for tech safety labour

In June 2026, Aapti Institute hosted a session at [All Things in Moderation](#) 2026 titled “Producing precarity: Labour conditions in content moderation”. It consisted of a short presentation on the precarious nature of content moderation and safety data work, a panel comprising two former content moderators and a researcher, followed by a concluding open discussion with participants. The intention was to discuss the labour costs and problems of sectors like content moderation for workers in the Global South specifically, and to collectively think about the principles or changes needed.

Context

Human labour is frequently responsible for tech products’ safety and usefulness. From content moderation to data work, humans do valuable work. However, workers in these roles have been found to also face precarious working conditions on multiple occasions. Issues like the pace of work and the personal risk involved, make it essential for the industry to change its practices for policy to intervene.

Our session sought to offer some discussions on content moderation and to foster discussion about principles or actions to pursue across the sector.

Takeaways from the session

The panel brought together two former content moderators and a researcher, and the discussion across all three contributions pointed to a number of consistent and significant findings.

1. **Mental health support is structurally insufficient:** Across both moderator testimonies, a clear pattern emerged: mental health provisions in content moderation are typically outsourced alongside the work itself, placed with vendors whose primary incentive is cost reduction. This results in support being delivered by personnel who are underqualified to address the conditions workers face. In some cases, wellness staff operate in proximity to management, raising concerns about confidentiality. One panellist described inviting a coach to observe the content being moderated, the coach left after a single session and did not return. The inadequacy of support structures reflects a broader failure to treat psychological harm as an occupational health obligation.
2. **NDAs function as instruments as isolation:** Non-disclosure agreements in this sector extend beyond professional confidentiality. Workers are contractually prevented from discussing the nature of their work with family members, friends

or partners. This forecloses informal support networks and leaves workers unable to seek help outside of formal channels that are themselves inadequate.

3. **The mechanics of moderation produce compounding pressure:** Workers operate against daily productivity targets, in some cases engaging with 1,500 pieces of content per day alongside accuracy thresholds of 90–95%, applied against detailed and frequently changing policy frameworks. Content is reviewed in isolation, without contextual information. Decisions must be rapid, accurate, and individually justifiable. These conditions concentrate pressure on workers in ways that directly increase exposure to harmful material and leave limited margin for error.
4. **Workers are contributing to AI development without disclosure or compensation:** Moderators have in some cases been training automated systems on the backend without being informed that this was occurring and without additional compensation.
5. **Pay disparities across geographies are significant and unaddressed:** Workers performing the same tasks in different jurisdictions receive vastly different compensation, with no accountability mechanism to address this imbalance. For example, content moderators in Australia are paid significantly more compared to content moderators who sit as outsourced labour in regions such as Africa, Southeast Asia and South Asia. This reproduces broader inequalities within global outsourcing models rather than differences in the nature or value of the work.

Action Points:

The following points were raised across the panel and the open discussion as areas requiring attention:

1. **Reframe client policies to include worker welfare:** Moderation is currently governed by what the client specifies. Platform policies should also include enforceable provisions on working conditions, mental health standards, and disclosure obligations.
2. **Mandate transparency from platforms:** The scale at which content moderation labour is performed remains unknown. As a foundational action, platforms should be required to publish the number of people performing moderation work across their trust and safety operations, distinguishing between in-house teams and outsourced workers across the supply chain.

Alongside this, transparency on the total expenditure allocated to trust and safety and what proportion goes to human labour versus technology would allow regulators, researchers, and civil society to assess whether investment is proportionate to the scale and risk of the work.

3. **Conduct Labour Mapping:** A comprehensive mapping of who is performing content moderation and data work globally - across vendors, geographies and platforms is a prerequisite for effective governance.
4. **Reform hiring and onboarding practices:** Resilience and psychological assessments should be conducted prior to exposure to harmful content. Job descriptions should accurately reflect the nature of the work.
5. **Address career progression and long-term wellbeing:** Content moderation is currently structured as a short-term, high-intensity role with limited progression. Developing certification frameworks, career pathways, and portability of skills would begin to address the structural undervaluation of this work.
6. **Foster collaboration and co-designing:** It would be beneficial to have all decision-making stakeholders and those affected by the decisions to work together to understand the issues in the sector and craft solutions in a collaborative manner.

These recommendations reflect key themes that emerged throughout the panel and open discussion. While they are not exhaustive, they accentuate several practical areas where further attention, collaboration, and research could help strengthen protections and support for content moderators.

We extend our sincere gratitude to the panellists for their participation and for the candor with which they shared their experiences and knowledge. Their contributions lent the discussion a depth and significance, and we are deeply appreciative of their willingness to engage with these issues in a public forum.

We also wish to thank the ATiM team for their support and for maintaining a conference that centres the experiences of those working in content moderation. We are equally grateful to the attendees whose questions and comments enriched the open discussion and highlighted the breadth of expertise present in this community.

This session represents a small share to a much larger, ongoing and necessary conversation. We hope it serves as a meaningful continuation of the important work our panellists and the wider ATiM community are already doing.

If you would like to reach out to us at Aapti, we can be reached over email via sreya@aapt.in or ritvik@aapti.in.