

RECOURSE: COMPANY-ESTABLISHED REVIEW BODIES

The following note explores elements related to the potential creation by companies of mechanisms to provide an independent appeal of their content restriction decisions made on the basis of their community guidelines. It is understood as a company-specific instrument¹ with binding authority at the third level of a decision-making escalation path following initial first instance decisions and reconsideration².

As detailed below, analogies with national Supreme Courts (or equivalent) have some validity but can only be pushed so far, given very significant differences in situations. Other inspirations might be also relevant and inventiveness is required in this radically new transnational environment.

The following preliminary questions could help structure "what if" discussions on the creation of such a mechanism:

COMPETENCE	DUE PROCESS	BODY	OTHER
<ul style="list-style-type: none"> • Topics covered • Normative reference • Initial source (AI/notices) • Cases filtering ("cert") • Mandate focus/limitation • Applicants • Remedies 	<ul style="list-style-type: none"> • Limited steps/duration • Written/oral procedure • Adversarial process • Role of third parties • Decision-making • Production of rationale • Dissenting opinions • Expedited mechanisms • Transparency • Suspensive procedure 	<ul style="list-style-type: none"> • Size • Composition • Members profiles • Designation • Mandate duration • Meeting frequency • Independence • Secretariat support • Funding 	<ul style="list-style-type: none"> • Charter • Name • Advisory role(s) • Geographic scope • Thematic chambers • Mutualization • Electronic tools • Liability protection • Jurisprudence coherence

The elements in the table above are briefly detailed below:

1. **Competence**

- a. Topics covered:** Community Guidelines cover diverse topics with different volumes of content restrictions and levels of automatic detection³. To keep the volume of expected appeal requests manageable, should such a mechanism initially be open only for certain topics? Would an option in that regard be to focus on issues (e.g. hate speech and bullying) where the impact on freedom of expression and the need for nuance are maximum, while the number of initial actions is relatively smaller (see annex)?
- b. Normative reference:** Community Guidelines would clearly constitute the primary normative source. Should however the Charter of such a body also reference other sources, such as international law (e.g. human rights principles and specific conventions) or even national laws (particularly if the body's remit also covers at some point requests from public authorities on the basis of national law)?
- c. Initial detection:** Content restriction decisions are taken on the basis of 1) Artificial Intelligence detection, or 2) flagging by users or notices by public authorities. Appeals in the first case only involve the posting user and the company, while the second case creates a tripartite interaction, with potential impact on the procedure. Should the envisaged mechanism only concern the first case, for the sake of simplicity, or cover both situations? Could this be established in phases?
- d. Cases filtering ("cert"):** Preventing the docket overload plaguing many high courts around the world is a fundamental success factor. The US Supreme Court only "grants cert" to 1,4 % of the annual submissions it receives (100 out of 7.000) and other jurisdictions have a similar practice. Case filtering is probably needed here but there is uncertainty as to the actual number of cases to be handled. Beyond weeding out clearly frivolous cases, does this require a rapid early selection mechanism? What latitude should the body have in selecting the cases it handles? Should there be specific expedited provisions to prioritize issues with a timeliness factor, such as urgency or tense local situations?
- e. Mandate focus/limitation:** Most Supreme Courts adopt or are subject to a restrictive approach in their case selection, in order for instance to focus on: conflicts between lower jurisdictions; clear challenges of interpretation of the law or a Constitution; or procedural dimensions of a lower judgment. Should a similar approach be applied here and documented in the body's Charter and online submission forms? Yet, it is important to note that in

the examples above, the ultimate review mechanism sits on top of two levels of lower independent courts and a large body of public jurisprudence while the exercise here starts *de novo*, and is expected to intervene directly as a follow up to a simple internal process of reconsideration.

f. Applicants: An ultimate independent review is envisaged only against restriction decisions made by the existing appeal/reconsideration stage and not against an initial decision (rule of exhaustion of previous avenues for recourse). It is clearly intended to be open for a user whose content has been restricted. Should this appeal also be open - and if yes, when - to notifiers whose requests for removal have been denied? In that case, should distinctions be made between public authorities and individual flaggers? And among the latter, between people directly targeted by the post and more general flaggers? Opening recourse to notifiers adds procedural complexities.

g. Remedies: What range of remedies can be ordered: mere reversal of the platform decision or also more granular and nuanced alternatives (e.g. technical, geographic scope, warnings...)? Can the body order the company to post a public correction?

2. Due process

To fully respect human rights, an independent review body must take inspiration from elaborate due process requirements developed in various nations for courts dealing with freedom of expression. However, the expected large volume of cases and the need to keep the process manageable call for some adaptation. This means, *inter alia*, making choices on the following elements:

a. Limited steps/duration: Rather than several iterative phases, should the procedure have a limited number of steps and/or duration? Can dedicated online formats for appeal help in that regard?

b. Written/oral procedure: Would the procedure be based exclusively on written briefs or on oral arguments as well? Would this vary in any way depending on cases?

c. Adversarial process: This procedure can be seen in two different ways. Either as arbitrating a dispute between the company and the user, or as reviewing a lower level decision. Would company representatives be a party in the procedure or only provide reasoning for the initial decision? Should individual notifiers directly affected by the posting under evaluation (if applicable) be part of the process?

- d. Third parties:** What is the possibility of their intervention in the procedure (e.g. legal representation, amicus provided by a supporting NGO or other parties) and conditions thereof?
- e. Decision-making:** What would be the majority rules for the body and any subsets of it?
- f. Production of rationale:** Producing a rationale for every decision is potentially burdensome but an important contribution in setting up a coherent jurisprudence, as it establishes precedent. Should this be implemented and if so, for all or only certain decisions (for instance in larger formations)?
- g. Dissenting opinions:** Can they be envisaged, and if so, under which conditions?
- h. Expedited mechanisms:** Irrespective of the overall body size, can most decisions be made by a limited number of members, keeping larger formations for more delicate cases? Likewise, can procedural guarantees vary according to the perceived importance or complexity of the case, with for instance a mere one-step written procedure for the simpler ones?
- i. Transparency:** What would be the level of publicity of deliberations and decisions? Should a repository of such decisions be put in place, and if so, by whom? Should some precautions be taken to preserve the rights of users, such as anonymization of decisions?

3. Body

- a. Size:** If a single review body is formed per company, determination of its appropriate size can be informed by an analysis of practices in Supreme Courts or equivalent not only in the US, but also in Europe (e.g. France, the UK, Germany, ...), India, Brazil and other countries, as well as some regional ones (e.g. ECHR). These bodies significantly vary in size: 9 for the US Supreme Court, 30 in India, 47 for the ECHR, with varying formations sizes. In light of the expected large number of cases, the size and diversity of user communities, and the absence of a vast network of lower courts, can a small number of members be viable? On the basis of practices in arbitration and ADR, should the establishment of a large roster of available arbiters or mediators to compose ad hoc panels be explored?
- b. Balanced composition:** Community Guidelines cover several topics, requiring diversified expertise. Also, a company's global geographic range means a diversity of scripts⁴, languages, cultural and political local contexts, that call for linguistic and local knowledge. How to ensure the geographic, stakeholder, gender, age, cultural and competence balances that will be key to allow nuanced decisions and establish legitimacy of such

a body? Should the composition be organized around specific constituencies? How can the interests of the user community be represented?

- c. Members profiles:** A spontaneous impulse could be to mainly call on the expertise of former judges or lawyers. Yet, most are proficient in one particular body of national law and might keep a particular bias in that regard while the main body of norms to be enforced here is likely to be the Community Guidelines. How to ensure a broad diversity of professional profiles and experiences, privileging people with exposure to a variety of environments?
- d. Designation:** This may be one of the most delicate issues. Modalities of designation of superior court judges in countries are difficult to transpose here. A designation by company management alone - whatever the level - would probably not be perceived as fully legitimate. Yet, a full selection by the community itself raises many conceptual and operational challenges. What innovative mechanisms can be designed to enable the selection of people of high integrity, competence and dedication that will be seen by the community and the general public as forming a legitimate and trusted college? Should diverse sources of designation be combined? Should specific constituencies play a role and if yes, how to form them if it is at a global level?
- e. Mandate duration:** The lifelong mandate of US Supreme Court Justices is an outlier case and a limited mandate seems more appropriate. What would be its appropriate duration (2, 4, more years)? Should there be a limited number of renewals (1, 2, more)? Would partial rotation help avoid brutal changes in the composition of the group? Could a large group be established progressively (e.g. with 1/3 of the members every 2 years in the first 6 years, if such duration is retained)?
- f. Meeting frequency:** How frequently should the Body be in session? This should take into account the expected amount of cases, as a result of responses to the above questions of scope (in part 1) and procedure (in part 2).
- g. Independence:** This is a critical factor, whatever role the company might play in the designation of members. Given the expected volume of activity, should members be expected to be fully dedicated to this mission for the duration of their mandate or not? In any case, what should be the conflict of interest policies limiting their past or current activities, including their potential relations with the company?

h. Secretariat support: Such a body will need secretariat support to manage the process and conduct research. Could automation reduce the overall burden in comparison with existing judicial processes?

i. Funding: Will the financing of such a body be under the exclusive responsibility of the company? Or should there be other sources?

4. Other

a. Charter: A dedicated Charter for this independent review body will be necessary, detailing inter alia its mandate, normative reference basis, procedures, composition and mode of designation. How should it be developed and what role can the corresponding user community play in that regard?

b. Name: This note uses the expression "Independent Review Body" by default. Alternative names can be envisaged, such as Panel, Council, Committee, or equivalent. What could be a proper name, given that the term "Independent Review" has a strong benefit in terms of clarity.

c. Advisory role(s): In addition to the appellate role on individual decisions envisaged above, should the following additional advisory roles could also be envisaged for such a body:

- On a case by case basis early on, upon spontaneous request by the company in difficult or sensitive situations, even before a decision is made or the user is notified,
- In a more general way, to provide guidance on best practices and refining of Community Guidelines, on the basis the cases it handles or some that would be shared by the company.

In the first case and maybe also in the second one, would the company have the option to either follow the advice/recommendation (without any further justification), or not (in which case it may have to provide an explanation to the body, to help refine its jurisprudence)?

d. Thematic chambers: Many jurisdictions around the world have specialized chambers for different topics. Should in due time different sub-groupings of the body be formed, according to members' competences or personal preferences, to specialize in particular types of cases?

e. Geographic scope: Likewise, in order to ensure maximum understanding of local context, cultures and legal frameworks, should sub-groupings of members - in due time - compose specialized chambers, for instance on a script or linguistic basis (rather than a purely geographic or national one)? How to maintain cultural diversity in such groupings to preserve some

jurisprudential coherence? Can such independent review bodies alternatively be created at national level?

f. Mutualization: Implementation of an Independent Review Body by every single company might be difficult, especially for small ones. Could some groupings be envisaged among several companies?

g. Electronic tools: Like in the general moderation, significant automation of this independent review process can be achieved to manage the workflow of a large number of cases. Can innovative mechanisms be explored to enable collegial decision-making among people likely to be distributed in various locations around the world?

h. Liability protection: Would the establishment of such an Independent Review body impact the liability regime of the corresponding company?

i. Jurisprudence coherence: How to ensure compatibility between the decisions of a diversity of Independent Review Bodies from different companies?

NOTE: This note purposefully focuses mainly on decisions taken on the basis of AI detection. Independent review of decisions taken on the basis of notification raise additional procedural challenges and should also be envisaged in particular in two cases: flagging by individuals directly targeted or impacted by specific posts and notices by public authorities. This would require additional procedural steps and more analysis.