Comments on Facebook’s draft Charter for an independent review board

Bonavero Report No. 2/2019
6 May 2019

Stefan Theil and Kate O’Regan
EXECUTIVE SUMMARY

Facebook’s draft Charter proposes that the primary mission of the independent board is to review individual content moderation decisions taken by the company (the ‘appeal model’). Our Report suggests three viable mechanisms to select decisions from the vast amount of reported content: (1) Facebook could select decisions for review, (2) the secretariat of the independent board could fulfil this task; or (3) the selection could be outsourced to accredited third party organisations. We also believe a combination of all three mechanisms is viable.

Apart from the appeal model, there are two further models to consider. The independent board could monitor outcomes and address systemic concerns through a general audit of the Facebook content moderation system (the ‘audit model’). This model focuses on the quality, consistency and timeliness of decisions, as well as evaluating the training, development and support of content moderators and monitoring of algorithmic decisions. Alternatively, the independent board could review and oversee the development of community standards and associated policies and guidelines based on international human rights norms (the ‘normative review model’). Adopting a combination of these models is advisable.

In any case, international human rights norms should be the primary standard of review. They give specific content to the values that inform Facebook’s philosophical approach to content moderation, frame and guide the inevitable balancing of competing interests, and ensure coherent decisions that can withstand external scrutiny. The independent board needs suitable powers to enforce its decisions, including requiring Facebook (1) to reinstate or remove content, profiles, groups and pages, (2) to revise the content moderation system and procedures, and (3) to amend the community standards as well as associated content moderation policies and guidelines.

With respect to institutional structure, the Charter must secure the independence and transparency of the independent board. Financial independence is perhaps best achieved through a trust endowed by Facebook and managed by independent trustees who could also serve as an Oversight Committee. The Oversight Committee would supervise the filling of any vacancies of the membership, deal with complaints against members, and monitor the activities of the independent board. In order to ensure transparency and legitimacy, the independent board must publish all its decisions, preferably in several languages and ensure that they are intelligible even where redaction of personal details is necessary. Beyond that, the establishment of a well-staffed secretariat is crucial, and the Charter should lay down core principles that govern the procedures of the independent board.

The Charter should further regulate the composition, terms of office, selection, and criteria for appointment of members of the independent board. The number of members will depend on the models adopted and we propose a single non-renewable term of office of between six and nine years. The actual and perceived independence of the membership will be crucial. Facebook could (1) select the initial members directly, and going forward the independent board will fill vacancies; (2) appoint an Oversight Committee that will set the criteria and oversee the selection process; and (3) select members by lot from among its users through a transparent process with public selection criteria. Any selection criteria employed should ensure that the board members represent and reflect the global community of Facebook users: gender, ethnic, cultural and linguistic diversity should be ensured. In addition, members must have appropriate expertise, experience, integrity and be independently minded.
Facebook’s Draft Charter for an Independent Review Board

I. Introduction

Facebook has invited comments from experts and civil society on its draft charter ‘An oversight Board for Content Decisions’ (the ‘draft Charter’).1 The draft Charter lays out Facebook’s vision for an external board (the ‘independent board’) composed of independent experts ‘who will review Facebook’s most challenging content decisions’. According to the draft Charter, the independent board will be able to reverse content decisions by Facebook and influence the development of the content moderation policy. Facebook identifies the purpose of the independent board as providing oversight of content moderation decisions and to make Facebook more accountable. The draft Charter briefly outlines the envisioned structure, scope and authority of the independent board and raises questions as a starting point of discussion.

The following Comments on Facebook’s draft Charter for an Independent External Review Board. (‘the Report’) emerged from discussions at a Roundtable Workshop held at St. Anne’s College, University of Oxford on 1 March 2019 (the ‘Roundtable Workshop’). The Bonavero Institute of Human Rights is a dedicated institute within the Faculty of Law at the University of Oxford, housed in a new building at Mansfield College: its core mission is to undertake world-class research in the field of human rights law and foster public engagement in human rights issues beyond the academy.

While this Report addresses many of the issues raised in the draft Charter, we have followed a different structure which we believe emphasises the central importance of defining the mission of the independent board. From this, many further questions follow that require careful and nuanced consideration. The Report is thus intended as a contribution to ongoing global debates about both content moderation and tech regulation: it advocates an institutional design that fosters legitimacy, transparency and accountability of social media platforms.

The Report focuses in turn on three major areas of debate: (1) the mission of the independent board what issues it can and should address and practical concerns relating to the competing models; (2) the standards of review the independent board ought to apply in fulfilling its mission, and the corresponding implications for Facebook and its content moderation system; and finally (3) the institutional design of the independent board, how it will secure its independence, determine its membership, and the procedures employed. The Report draws on the insights and comments presented at the Roundtable Workshop, but it does not purport to comprehensively represent and incorporate all expressed views.

II. Mission

Defining a mission and determining an appropriate model for the independent board is crucial to the questions surrounding the standards of review and institutional structure. It is entirely possible and, in our view, desirable for the independent board to adopt all three suggested models (appeal, audit and normative review). The independent board would thus address a range of different concerns with the current content moderation system, while avoiding the drawbacks of focusing exclusively on any one model.

1. Appeal model

The draft Charter currently outlines an independent board engaged primarily with the review of content moderation decisions in ‘important and disputed’ cases.\(^2\) To the extent that Facebook intends to adopt a model focused on the review of individual cases (the ‘appeal model’), a core problem is the process by which decisions are selected for review from the vast number of internal appeals. Facebook currently receives millions of reports per day and approximately 70,000 decisions are considered by the internal Facebook appeal process every month.\(^3\) As the draft Charter foresees at most a few hundred members for the independent board working on a part-time basis, it will not be possible to review every internal Facebook decision.

Three different mechanisms could be adopted, and each has advantages and disadvantages. The first mechanism would empower Facebook to refer cases to the independent board, as suggested in the draft Charter.\(^4\) While likely an efficient mechanism, Facebook’s control of the docket would also raise concerns about legitimacy and transparency of the selection procedure for cases, which in turn might adversely affect the perceived independence and usefulness of the board.

A second mechanism is for the independent board to select cases for review itself. In practice, this task would fall to a secretariat of full-time administrative staff.\(^5\) The selection criteria would need to be outlined in general terms in the Charter and specified by the independent board through transparent internal guidelines. The secretariat could for instance be required to identify systemic problems and select representative cases accordingly. This approach would ensure independence from Facebook and enhance the perceived legitimacy of the independent board. A difficulty with this approach is that the secretariat would require unhindered access to Facebook’s content moderation system, which might raise concerns over user privacy and data protection. Additionally, some deserving cases may not be selected for review, depending on the selection criteria employed.

A third mechanism would rely on third parties accredited by the independent board to submit cases. These third parties might be civil society organisations, user groups, state-established bodies such as ombudspersons, or regulatory bodies. The independent board would have to

---


\(^3\) These figures emerged in debate during the Roundtable Workshop on the draft Charter, St. Anne’s College, University of Oxford in 1 March 2019; Chatham House rule precludes identifying the source of these figures.

\(^4\) Draft Charter: An oversight board for content decisions, p.3.

determine the criteria for accreditation and likely define an upper limit of accredited third parties. Third parties can then raise either individual appeals, groups of appeals with common features or challenge perceived systemic issues. A key challenge with this option is the accreditation process, which must be fair and perceived to be legitimate. The mechanism may also raise difficulties in relation to data sharing and privacy, particularly if the third parties seek to bring cases without the authorisation of affected users. All three mechanisms have drawbacks and are premised on the independent board adopting an appeal model as suggested in the draft Charter.

Permitting Facebook to manage the docket may be efficient, but it raises concerns with perceived independence and legitimacy of the independent board. Allowing the independent board to select cases itself is a better model as it will alleviate these concerns, but it will inevitably leave some deserving cases unaddressed. Finally, leaving appeals to accredited third parties would outsource the selection process, but risks that systemic issues are overlooked, and deserving cases may not be considered. All selection mechanisms under the appeal model are likely to lead to dissatisfaction especially if the quality of the Facebook internal appeal system is considered unfair, inconsistent or unreliable. These drawbacks could perhaps be mitigated if all three selection mechanisms were adopted, and worked cumulatively, thereby providing different routes to the independent board.

2. Audit model

The independent board could alternatively or additionally address the root causes of the case volume by auditing Facebook’s content moderation system (the ‘audit model’). This audit model requires the independent board to monitor and evaluate the quality of content moderation decisions based on systemic outcomes as opposed to arbitrating individual decisions. Facebook CEO Mark Zuckerberg recently suggested such a role for the independent board in a Washington Post op-ed. In our view the audit model could address three areas of concern relating to the content moderation system that are not captured by the appeal model.

First, the audit model provides independent monitoring of the effective and timely enforcement of content moderation standards and policies. This would allay concerns over conflicts of interest. Some commentators have suggested that delays in content moderation might be driven by Facebook’s business model, which thrives on consistent and long-term user engagement: the more time users spend on Facebook, the more complete the data and the more attractive the overall dataset to advertisers, who pay for precise data and targeting of user groups. Harmful or shocking content draws users and increases time spent on the platform. Independent monitoring and systemic improvements of the content moderation system that encourages the timely enforcement of community standards might address these concerns.

---


Secondly, the audit model can ensure independent oversight of the quality, consistency and timely enforcement of content moderation decisions, whether conducted by humans or algorithms. Despite the advances achieved with algorithmic flagging and content moderation decisions, their design and operation, as well as the Facebook internal appeal process is prone to systemic biases, mistakes and design flaws. Hence, independent oversight of the quality of the content moderation system remains essential even if significantly more decisions are automated in the future.

Thirdly, an audit model could oversee the content moderation infrastructure, such as the number of content moderators, their training and requisite levels of expertise, including linguistic and cultural expertise, as well as monitor terms and conditions on which content moderators are employed and arrangements to ensure their well-being. These issues are all of major concern to European regulators, as reflected notably in the German Network Enforcement law, which expressly provides standards for the training and management of content moderators.  

3. Normative review model

Finally, there is a third possible model for the independent board that focuses on the long term development of the content moderation standards. The independent board could oversee and ensure that community standards and policies are drafted and developed consistently with international human rights norms (the ‘normative review model’). Under this normative review model, the independent board would periodically review the community standards, including the iterative changes made to those standards through Facebook’s ‘Content Standards Forum’, as well as related policies and guidelines to determine whether they are fit for purpose and in line with higher order norms and principles. Based on this review, the independent board would recommend and require changes to the community standards and their practical application so that the content moderation system complies with human rights norms. The independent board would guarantee that that normative hierarchy outlined in the next section is implemented and informs all content moderation decisions, policy changes and future initiatives.

---

III. STANDARD OF REVIEW AND IMPLEMENTATION

Regardless of the models adopted for the independent board, a clear and transparent hierarchy of norms should form the binding standard of review. At the apex of this hierarchy are international human rights norms, which provide the standard against which the entire content moderation system and all aspects of its decision-making process are evaluated. The interpretation and application of international human rights norms to individual decisions is supplemented by the Facebook’s community standards, as well as associated guidelines and policies. The Charter meanwhile complements the hierarchy of norms by focusing on the implementation powers and procedures of the independent board.

1. Human rights norms

Relevant international human rights norms would include the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights (ICCPR), the International Covenant on Economic, Social and Cultural Rights (ICESCR), as well as regional protection regimes such as the African Charter on Human and Peoples’ Rights (ACHPR), the American Convention on Human Rights (ACHR), the American Declaration on the Rights and Duties of Man and the European Convention on Human Rights (ECHR). The ICCPR will be of particular importance because it has been ratified by 172 states worldwide. This approach is broadly aligned with suggestions made by the UN special rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression David Kaye.9

This emphasis on international human rights norms is crucial for four reasons. First, international human rights are based on important moral commitments that give specific content to the values that Facebook cites in the draft Charter as informing its philosophical approach to content moderation: voice, safety, equity, dignity, equality and privacy.10 Secondly, the worldwide endorsement, particularly of the ICCPR means that it provides a legitimate foundation for the principles of a global social media platform. Thirdly, while human rights are not self-executing and require further specification for practical application, they nonetheless guide and frame the balancing of competing interests. Human rights norms offer a well-established and widely accepted mechanism to balance competing interests of users, platform operators and the public interest. International human rights norms accept, for instance, that freedom of expression is subject to proportionate limitations, particularly in the interests of combatting hate speech. Fourthly, adopting international human rights standards makes it easier to offer a coherent and principled defence should the decisions of the independent board be subject to external scrutiny, for instance in domestic and international courts. In this vein, the independent board should also take account of the decisions issued by the relevant treaty bodies and courts, as well as authoritative commentaries and scholarly publications.

9 David Kaye, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (UN A/HRC/38/35, 2019).

10 Draft Charter: An oversight board for content decisions, p.3.
2. Community standards and local laws

The Facebook community standards should form the second and supplemental tier of norms applied by the independent board. The draft Charter accepts that decisions of the independent board will be binding on Facebook in individual cases and may require reforms of the content moderation system as a whole.\textsuperscript{11} We believe it is crucial that the independent board has the power to bindingly require Facebook to enact changes to the content moderation system, especially if it adopts the normative review model.

Difficulties will arise when local laws are inconsistent with international human rights norms, a particular challenge for a global social media platform. The same reasons that ultimately justify the primacy of international human rights norms also explain why the independent board should be reluctant to permit departures from them. We take issue with the assertion that ‘[t]he board will not decide cases where reversing Facebook's decision would violate the law.’\textsuperscript{12} The independent board should review any decision it chooses, including those where content moderation decisions are taken in order to comply with local laws. Where such decisions clearly and unreasonably infringe international human rights norms, the independent board should make that clear in its decision. Facebook must then decide how to proceed, given the risk that it will be seen to be complicit in the suppression of human rights.

3. Implementation

The Charter must equip the independent board with suitable remedial powers to enforce its decisions. Under an appeal model the board should be empowered to reverse or require the removal of content, profiles, groups and pages. Under an audit model, the independent board should focus on systemic outcomes and hence be empowered to require Facebook to address weaknesses or flaws in the content moderation procedures, which might include revision of algorithmic design, setting time frames for content decisions, reviewing the remedial powers of content moderators, setting qualifications for their appointment as well as reviewing employment conditions, benefits, and training. The normative review model should be empowered to require amendment of the community standards, associated guidelines and policies as well as to set rules for the application of the community standards, and associated guidelines and policies. Regardless of the model, Facebook should have no discretion to refuse the implementation of decisions of the independent board.

The Charter should further specify whether and what kind of remedial orders the independent board can issue beyond requiring the deletion or reinstating of content. This could for instance entail financial compensation for users or public and prominent acknowledgment of wrongdoing. Facebook should further be obliged not to amend its community standards in a manner that contravenes a decision of the independent board or that frustrates required changes to the content moderation system. Crucially, the independent board must have the exclusive authority to determine whether Facebook has complied fully with its decisions and remedies.

\textsuperscript{11} Ibid, p.3.

\textsuperscript{12} Ibid, p.3.
IV. INSTITUTIONAL STRUCTURE

The key institutional features of the independent board must be specified in the Charter. The finalized Charter must secure first and foremost the institutional independence and transparency of the independent board. It should also regulate the composition, terms of office, selection, and criteria for appointment of its members. Beyond that, it should provide for the establishment of a secretariat, as well as core principles that should govern the working rules and procedures of the independent board.

More specific and practical considerations that flow from the structure and general principles outlined in the Charter may be autonomously determined by the independent board, preferably in a set of transparent and easily amendable working rules. In this Report, we focus on the Charter, leaving for the future a more comprehensive discussion of the working rules that should be adopted by the independent board.

1. Institutional independence

Institutional independence is crucial to the effective operation of the independent board. One of the keys to institutional independence is financial independence from Facebook and other outside funding sources. To that end, the independent board should be endowed with sufficient funding to ensure its proper and continuous operation. Likely costs that must be covered through the endowment include remuneration for its part-time members (including travel and accommodation), the staffing of a suitably sized secretariat that organises the day-to-day operation and coordinates with Facebook on information exchange (salaries, wages, institutional pension and healthcare plans) and any expenses related to the operation (office space, hardware, software) of the independent board, including expenses associated with the decision-making process (travel and accommodation expenses for expert witnesses, translation costs, and related costs such as legal aid and a litigation budget).

The necessary financial independence will perhaps best be provided by establishing a trust endowed by Facebook and managed by a group of independent trustees. Independent trustees could serve as the committee that selects the membership of the independent board. In this way, they would form an Oversight Committee that would manage the endowment, oversee the filling of any vacancies of the membership, deal with complaints against members, monitor the activities and safeguard the independence of the independent board. The Oversight Committee would comment on these developments through annual reports and be empowered under the Charter to take action as necessary. The Oversight Committee members would be required to meet the same requirements in terms of reputation, expertise and independence that apply to the membership of the independent board.

2. Transparency

The Charter should ensure the transparency and legitimacy of the independent board’s decisions. To that end, the independent board must publish all its decisions and ensure that they are intelligible to outsiders, even where redaction of personal details is necessary. Given the global community of users, the independent board should consider publication of its decisions in several languages. The independent board should take note of the needs of individuals with
various forms of disadvantage, including the visually and mentally impaired. The independent board must have adequate resources to defend its reasoning when its decisions are challenged in domestic and regional courts. The Charter should specify that the independent board should have at least two and up to five working languages, and should be equipped to receive complaints in further languages to ensure equal worldwide access and fairness.

3. Composition of board
The number of members of the independent board will depend on its mission and the models adopted. The greater the number of members, the more difficult it will be to develop a coherent and consistent approach to the decisions. However, the size of the board will also need to be sufficiently large to perform its function under the different models previously discussed. In addition, the Charter should regulate the size of panels for decision that is the number of members of the independent board required to issue a decision. It might be possible to provide a panel of three for ordinary decisions, with a specially constituted grand panel of nine or even fifteen members for important decisions, which might include the review of earlier decisions made by the independent board.

4. Terms of office
The terms of office of the members of the independent board should be non-renewable to enhance their perceived independence. However, terms should not be too short as it will take time for members of the board to become acquainted with procedures and decision-making process. Accordingly, we propose a single term of between six and nine years, which is not renewable.

5. Selection of members
The selection of members for the independent board must be regulated by the Charter. The actual and perceived independence and diversity of the membership will be crucial in securing the legitimacy of the independent board. Three options appear viable, but we accept that all have disadvantages.

Facebook could select the initial members of the board directly based on a public ‘review of qualifications’, as has been suggested in the draft Charter. Going forward, the draft Charter proposes that members of the independent board will select future members. The advantage of this approach is efficiency, but the perceived independence of the members will be damaged if Facebook is seen as too involved in their appointment. Moreover, it is undesirable for an independent board performing a crucial public function to be entirely responsible for nominating its membership. For the independent board to be seen as a self-perpetuating body will undermine its legitimacy.

A second option would be for Facebook to call publicly for nominations to a selection committee that will independently set the criteria and oversee the selection process of members.

\[13 \text{ Ibid, p.2.}\]
The selection committee should be composed of individuals that enjoy high levels of trust in their respective fields. The selection committee could call for nominations and conduct a process of testing and interview of potential members as in some modern judicial selection procedures. A selection committee that followed a transparent and public process would enhance the perceived legitimacy and independence of the independent board and could serve as an Oversight Board in the manner outlined previously.

Finally, the initial members of the independent board could be drawn by lot from active users of Facebook, through a transparent process with public selection criteria, including age, expertise and independence. This would add an element of randomness and avoid concerns over an excessively technocratic membership of the independent board. However, there are substantial concerns about such a process. The time commitment expected by randomly selected users would be significant and even though there are some analogies to the selection of jurors in criminal trials, these are generally concerned with questions of fact on a one-off basis, and not repeated, long-term substantive decision-making. A closer analogy might be found with citizen’s assemblies, which make recommendations and, in some cases, decide contentious questions following in-depth briefings and consultation with subject matter experts. None of these concerns exclude the possibility of selecting at least part of the membership of the independent board through a user lottery.

It might also be possible to accommodate several selection procedures: some members could be selected by an independent selection committee, while others are drawn by lot from Facebook users. In any case, all future vacancies should be filled through the same selection procedure that applied to the initial appointment.

6. **Criteria for appointment**

Any selection criteria employed should ensure that the board members represent and reflect the global community of Facebook users: gender, ethnic, cultural and linguistic diversity should be ensured to the extent possible, in proportion to the size of the cohort.\(^\text{14}\) In addition, there should be requirements of expertise, experience, integrity and independent-mindedness on the part of all members. Once the initial members are selected, they should choose from amongst themselves a President and Vice-President who are jointly accountable to the members of the independent board and empowered to speak on their behalf.

7. **Secretariat**

The independent board must be supported in its operation by a secretariat. It should be staffed by suitably qualified individuals hired directly by the independent board through an open and global application process. Facebook should not be involved in the hiring and firing process and the independent board should have its own separate human resources department.

\(^\text{14}\) Ibid, p.1.
8. Founding principles: procedural fairness and equality of arms

The Charter should further provide that the working rules and procedures adopted by the independent board must be consistent with the principles of procedural fairness, and transparency. Where lawyers are appearing before the independent board, the Charter must ensure that there is appropriate equality of arms. Users should be permitted to be legally represented and in appropriate cases legal assistance should be afforded to them to enable the defence of their claims.

9. Working Rules

Once these questions of institutional design and general principles have been settled in the Charter, the independent board should address more specific procedural questions in its working rules. These will be published and updated periodically as the independent board adapts and develops its working methods. The independent board would likely want to consider some of the following questions.

While basic procedural fairness is enshrined in the Charter, the independent board should give consideration to more specific procedural rules: who has standing to bring cases, are interveners permitted to submit amicus briefs, what are the standards of proof, are proceedings conducted in an inquisitorial or adversarial style, how does the independent board reach decisions, presumably by a majority, and will decisions permit dissenting opinions?

In the event that third parties are empowered to bring cases, the independent board will likely wish to consider collective action systems and how these can be implemented. Third parties should be empowered to bring claims on behalf of affected individuals that are affected by a systemic issue. The mechanisms by which individuals can join or withdraw from an action must be simple, clear, and accessible. For any such system, it is important to ensure democratic participation of affected users and transparency in the decision-making of the third-party organisations by providing forums for class deliberation and input.
APPENDIX

Participants of the Roundtable Workshop to consider Facebook’s proposal to introduce an independent external review board of its content decisions

Friday 1st March 2019 09h30 – 11h30

1. Kira Allmann, Oxford (kira.allmann@csls.ox.ac.uk)
2. Sophie Boyron, Birmingham (s.n.y.boyron@bham.ac.uk)
3. Oliver Butler, Oxford (oliver.butler@wadham.ox.ac.uk)
4. Paul Craig, Oxford (paul.craig@law.ox.ac.uk)
5. Horst Eidenmuller, Oxford (horst.eidenmueller@law.ox.ac.uk)
6. Timothy Garton Ash, Oxford (timothy.gartonash@sant.ox.ac.uk)
7. Robert Gorwa, Oxford (robert.gorwa@politics.ox.ac.uk)
8. Jeff King, University College London (jeff.king@ucl.ac.uk)
9. Eleanor Marchant, Oxford (eleanor.marchant@csls.ox.ac.uk)
10. Michael Molavi, Oxford (michael.molavi@law.ox.ac.uk)
11. Helen Mountfield QC (helen.mountfield@law.ox.ac.uk)
12. Ravi Naik, ITN Solicitors
13. Adam Perry, Oxford (adam.perry@law.ox.ac.uk)
14. Nate Persily, Stanford (npersily@law.stanford.edu)
15. Gavin Phillipson, Durham
16. Andy O’Connell, Facebook
17. Catherine O’Regan, Oxford (catherine.oregan@law.ox.ac.uk)
18. Jake Rowbottom, Oxford (jacob.rowbottom@univ.ox.ac.uk)
19. Stefan Theil, Oxford (stefan.theil@law.ox.ac.uk)
20. Joe Tomlinson, University College London (j.tomlinson@publiclawproject.org.uk)
21. Sandra Wachter, Oxford (sandra.wachter@oii.ox.ac.uk)

Other individuals present declined to be included on the list of participants.