

The Acquitas Principles
on
Online Due Process

1	<i>Scope of Principles</i>	2
2	<i>General Principles of Platform Decision-Making Concerning User-Generated Content</i>	4
3	<i>Platform Decision-Making Concerning Online Criminal Content – Measures to Address Under-blocking</i>	6
4	<i>Platform Decision-Making Concerning User-Generated Content – Measures to Address Over-blocking</i>	7
4.1	Overview	7
4.2	Appeals Panel	9
4.3	Automated Review	10
4.4	Human Review	10
4.5	Algorithmic Enforcement	11
4.6	Full Disclosure and Transparency Principles	11
5	<i>Digital Literacy</i>	12

1 Scope of Principles

Human rights and fundamental freedoms must be upheld everywhere; they are no less important in the online environment than they are offline. Generally, governments are responsible for the safeguarding of human rights and fundamental freedoms; however, it is not only the activities of governments that impact human rights. On the internet, content platforms (“platforms”) create spaces, structures and processes for interaction, education, work and commerce. They therefore play a crucial role in shaping the online environment and an individual’s ability to enjoy their rights and freedoms. That role is even more important when a platform enjoys a dominant position in a given market. Notwithstanding the modern public square role of platforms, national and international laws and national and international courts have not extended the responsibility for the maintenance of human rights to platforms. Some supreme courts have, however, highlighted the public square or freedom of expression functions of platforms.

Accordingly, platforms should take an active role in preserving human rights and fundamental freedoms in the online world. The responsibility for the safeguarding of human rights and fundamental freedoms on the internet may be extended to platforms through laws, including international laws. However, even when such laws do not exist, platforms should take active steps to preserve human rights – whether as a part of corporate social responsibility or as a matter of business ethics. Some platforms have already recognized the effects of their activity on human rights and fundamental freedoms, and all platforms should convert that recognition into actual steps in furtherance of individuals’ rights and freedoms on the internet.

Processes that are consistent with due process principles are crucially important for the maintaining of procedural fairness and substantive justice, and due process principles should permeate all activities of platforms. Some activities of platforms are regulated by law or by the courts and agency decisions, which should incorporate due process principles. Other activities of platforms might not be regulated by law or by the courts and agency decisions, and in such cases, it is up to the platforms to implement processes that safeguard due process principles.

Platforms should apply due process principles to every decision that affects their users and other parties. The importance of the maintenance of the principles has become increasingly significant as platforms have engaged in what has been perceived as the under-blocking and over-blocking of online content. Due process principles should assist in mitigating under- and over-blocking, and platforms should therefore apply due process principles when they moderate content, curate private data, make online disclosures, publish transparency statements, and conduct other online activities. The establishment of a proper balance among different fundamental rights, including but not limited to freedom of expression, protection of personal data and protection of intellectual property rights – as well as upholding the rule of law in preventing crime online – all depend on the proper application of due process in the online environment. Platforms must play an integral role in the maintenance of due process online.

Public discourse, freedom of expression, and global online crime prevention are at present under extreme pressure, from both significant over-blocking and under-blocking by platforms. At one end of the spectrum, user-generated content is often subject to algorithmic content identification and blocked even in the absence of any legal violations. At the other end, clearly criminal content such as child sexual abuse and human trafficking is not sufficiently blocked throughout the world. These extremes in content blocking may be caused by inadequate and unbalanced decision-making processes which platforms apply. Due process principles, including procedural fairness and natural justice, should guide the design and implementation of platform processes so that human rights, fundamental freedoms and the rule of law are restored and maintained in the online environment.

Under these circumstances, concerted efforts should be undertaken to coordinate principles on an international scale for platforms' decision-making processes in serious, flagrant or egregious cases. Current areas of global concern are, for example, child sexual abuse and human trafficking online and similarly egregious crimes against humanity. Equally, tackling egregious violations of freedom of expression related to user-generated content require coordination. Such coordination, in the interest of legal certainty and justice, should preferably lead to the development of standardised procedures which are clear, accessible and user-friendly. Standardisation should also aim at improving transparency and

accountability by producing reliable data to facilitate third-party scrutiny by regulators, NGOs and other stakeholders.

Platforms cover a broad spectrum in terms of their financial and technical resources and their capabilities and functionalities. So compliance with due process requirements should be assessed with a view to proportionality and reasonableness, taking into account the specific circumstances in any particular context.

2 General Principles of Platform Decision-Making Concerning User-Generated Content

Due process should be reflected in platforms' decision-making concerning user-generated content and include the following principles.

1. A fair and transparent review within a reasonable time and by an independent and impartial decision-maker
2. Proper prior notification to the parties
3. An opportunity for the parties to respond and present evidence
4. The right to legal representation
5. The right to appeal to an internal appeals panel, alternative dispute resolution panel or competent court, and
6. The right to receive a decision in writing (or other perceivable form) and in a preservable format which clearly articulates the reason for the decision.

As part of these due process principles, platforms should provide clear information to users about the platforms' decision-making structures and procedures. Such information should include a simple, easy-to-understand notification to a user when content is taken down or blocked by a platform, together with clear reasons for the take-down and, when applicable, the legal basis for the take-down. Platforms should also provide a clear and simple outline for the process of challenging the take-down, whether through a counter-notice or other equivalent means.

Users should be given an opportunity to request timely follow-up actions and decisions by platforms. The rules governing users' actions (as well as rules and timelines on notice and take-down, submission of evidence, appeals procedures and timelines for all decision-making processes) should be clear, publicly accessible, easy to understand and applied in a timely manner.

Platforms should continuously evaluate whether their decision-making processes are designed and implemented adequately to reflect due process principles. The decision-making process must not be entirely automated; there should be provision for adequate and balanced input via human review, at least at the appeals stage. Platforms should provide for proper, adequate and professional qualifications and safety training for reviewers and other decision-makers.

Platforms should safeguard user privacy, including the principles of lawfulness, fairness and transparency; limitation on purposes for data collection, processing and storage; data minimization; accuracy of data; data storage limits; integrity, security and confidentiality; and accountability. Where appropriate, and even if not so required by law, platforms should make information available to users about the users' data, and provide for data rectification and deletion, data portability and control of legacy accounts.

Platforms should design their processes so that they are available only to notifiers, such as law enforcement authorities, who themselves implement fair internal processes for such notifications. To be effective, notifiers must satisfy all procedural and substantive requirements. Notifiers must play their part by implementing adequate internal processes, and this applies also to government agencies. For instance, a law enforcement agency using an expedited take-down process which is not subject to judicial or independent third-party oversight should implement internal due process safeguards to prevent the risk of false positives. Preventing false positives is the responsibility of notifiers and platforms.

Independent oversight is important to ensuring compliance with due process principles. Platforms' decision-making processes should not prevent a user from seeking judicial review by a court or an independent alternative dispute resolution body, whose procedures should safeguard due process principles.

3 Platform Decision-Making Concerning Online Criminal Content – Measures to Address Under-blocking

Platform decision-making concerning criminal content should include the following platform review and judicial review processes.

Platforms must comply with court and agency orders that require take-down of criminal content. Additionally and to the extent permitted by law, platforms should also respond to other notifications of criminal content. For these notifications, platforms should establish prompt, transparent and efficient procedures for submitting and reviewing criminal content notifications and for implementing effective remedies.

Unless provided for by law, platforms should prescribe the content and form of a proper notification together with:

1. Applicable deadlines
2. Timelines for follow-up by platforms
3. An appropriate format for the submission of evidence, and
4. A procedure for the removal of criminal content in a timely manner or, in some instances where warranted, on an expedited basis.

Platform decision-making concerning the removal of criminal content may include or provide for the following platform review and judicial review processes.

1. In urgent circumstances, platforms should make provision for the expedited removal of (serious/flagrant/egregious) criminal content or content which involves the (serious/flagrant/egregious) violation of human rights and which cause imminent harm. Platforms should remove or disable access to such content (as soon as is practically possible and on an urgent basis) following receipt and review of the notification.

2. Expedited removal should follow an expedited review by the platform when so required by law, such as pursuant to court orders and/or notifications by law enforcement authorities, or when requested by trusted parties, such as other relevant government agencies or well-established relevant non-governmental organisations (NGOs).
3. In other circumstances, platforms should make provision for the removal of criminal content or content which involves the violation of human rights in a timely manner. Platforms should remove or disable access to such content as soon as is reasonably possible following receipt and review of the removal order or notification.
4. Due process must be maintained also at the source; therefore, prior to issuing a content removal notification, notifiers such as law enforcement authorities, government agencies and NGOs should establish that they have the appropriate legal basis, such as probable cause or reasonable grounds, to conclude that the activity being notified is illegal.
5. Platforms should limit or remove the ability to use expedited and other procedures in cases of notifiers who have consistently high false positive rates over a specified period of time and/or misuse the expedited procedures to suppress lawful expression. Cases when take-downs are mandated by law or court order and/or notification by a law enforcement agency should be excepted from this principle.

4 Platform Decision-Making Concerning User-Generated Content – Measures to Address Over-blocking

4.1 Overview

Platform decision-making concerning user-generated content should include the following platform review and judicial review processes.

Platforms must comply with laws and court and agency orders that require take-down of illegal content, such as copyright-infringing content. Additionally and to the extent permitted by law, platforms may moderate user-generated content either in accordance with their own standards or based on take-down complaints or take-down notifications (“complaints”). In these circumstances, platforms should provide for prompt, transparent and efficient review of complaints that users submit against take-downs of their content, and implement effective remedies, including content restoration (put-back).

Platforms should establish a clear, simple and easy-to-understand procedure for notifying users about take-down actions or complaints that the platforms receive. Platforms should prescribe the content of a proper complaint and counter-notice and establish procedures to handle unjustified take-down complaints.

Procedural matters prescribed by platforms should include:

1. Complaint and counter-notice deadlines
2. Timely follow-up actions by platforms
3. An opportunity for a complainant to submit evidence in the appropriate format, as specified by the platform, and within a reasonable time frame, and
4. Appropriate and quick appeals procedures, including an opportunity for the parties to be heard, resulting in a timely review decision.

Platforms should not make it unduly difficult to provide evidence, nor should a platform establish stricter requirements than those envisaged by law.

Platform decision-making concerning the removal of user-generated content may include or provide for the following platform review and judicial review processes.

1. Even when they are not required to do so by law, platforms should, where appropriate, establish a notice-and-take-down procedure which allows for the timely removal of law-violating content or content which involves the violation of a platform’s standards. Platforms should provide for prompt, transparent and efficient review of complaints. Where appropriate, platforms should remove or disable access to such content (as soon as is reasonably possible) following receipt and review of the complaint. In

appropriate circumstances, platforms should also provide for other effective remedies including but not limited to: notice-and-notice, content-labelling, warning, stay-down, suspension, counterspeech, account termination, de-indexing, unmasking, blocklisting and assigning strikes as an alternative to take-downs.

2. Platforms should notify users whose content is removed about the specific reasons for the removal. Platforms should also inform users, in a simple and clear manner, of procedures such as counter-notices which are available to appeal the removal of their content. Such procedures should be easy to understand, user-friendly and outlined in simple terms.
3. Platforms' and complainants' decision-making concerning user-generated content shall not, in principle, prevent users from making content available which is lawful. This includes content covered by freedom of expression, fair use, and other exceptions or limitations to the rights of others, as long as such content does not violate a platform's standards.
4. In order to avoid over-blocking and preserve human rights and fundamental freedoms, including freedom of expression and the freedom to conduct a business, governments should create the following legal environment. With the exception of (serious/flagrant/egregious) illegal or (serious/flagrant/egregious) infringing content, platforms should not be liable for keeping content online and available while an assessment of its legality is completed, unless applicable national law requires that platforms provisionally remove the allegedly infringing material before such assessment is completed.

4.2 Appeals Panel

Where a user or complainant files an appeal against a platform's decision to remove or keep available content, the decision should be reviewed by a competent, independent and impartial decision-maker, such as an appeals panel, in a timely manner. On appeal, the appeals panel should assess the compliance of the decision

with the platform's standards. Appeals panel members should have adequate legal and professional training, accreditation and independent standing.

The law may require a review by a competent alternative dispute resolution panel, administrative authority or a court – either in all cases or where the party or parties are not satisfied with the decision of the appeals panel. Even when the law does not so require, in cases when the party or parties are not satisfied with the decision of the appeals panel, platforms should facilitate parties' access to appeal to a competent alternative dispute resolution panel, an administrative authority or a court.

4.3 Automated Review

In response to the volume and velocity of complaints for removal, platforms may employ automated review as the first line of content review. Platforms may use automated removals, algorithmic take-downs and big data analytics to cope with the high volume of complaints.

However, where automated reviews and take-downs do not adequately process urgent or complex cases or where disputes arise which require human review, platforms should establish a clear and simple procedure for complainants to request human review.

4.4 Human Review

In cases which require more urgent review or involve complex issues, complaints for removal and appeals of removal decisions should be forwarded to a platform's human review team in a timely manner. Such cases should include those related to the appropriate weighing of rights and interests, in particular human rights and fundamental freedoms, and those where disputes have arisen over automated content removal decisions. The human review team should assess whether the removal has been correctly carried out according to the platforms' standards.

A clear and simple procedure should be made easily available for a timely appeal to an independent appeals panel in those cases where disputes have not been settled despite the review by a platform's human review team.

Human review team members and appeals panel members should be sufficiently trained, prepared and supported in their professional functions, particularly in cases when they are obliged to review distressing content.

4.5 Algorithmic Enforcement

Governments should carefully weigh the benefits and drawbacks of algorithmic enforcement of rights when deciding whether and to what extent the use of algorithmic enforcement tools should be mandated. Accountability, transparency, and non-discrimination requirements should apply to platforms which employ such tools, whether to comply with the law or voluntarily, including:

1. Algorithm, training data and decision-making logic transparency
2. The application of human-in-command principles
3. External audits
4. Liability and redress for AI-generated harm, and
5. Transparency reports.

4.6 Full Disclosure and Transparency Principles

Platforms should provide clear information to users about the platforms' standards and procedures. The criteria that the platforms use for removing content or for blocking content should be made clear in a transparent manner. Where a user is affected by a platform decision, platforms should inform users about the reasons for the decision. This should include which law and/or platform standards the content violated and how the manual or automated processes were used to identify the violation.

Platforms should also publish general information about their content moderation practices to enable regulators, government bodies, NGOs and other stakeholders to understand these processes and hold platforms and complainants accountable.

Additionally, in the interest of transparency, full disclosure by platforms should include regular publication of the following information:

1. Posts removed
2. Accounts banned / accounts suspended, and
3. The nature of the complaints made.

5 Digital Literacy

To facilitate users' awareness of and effective involvement in the procedures above, platforms should contribute towards efforts to improve online literacy. For example, platforms may raise awareness of due process mechanisms and assist users in learning how to use online tools to confirm the authenticity of products and offerings online and the accuracy of online content.

When designing policies and procedures to implement due process principles, platforms should take into account any potential barriers to access that users may experience, including varying degrees of online literacy of their users. Platforms should ensure that such policies and procedures are accessible to all users, including to members of systematically excluded groups and users with disabilities.