Content Regulation in the Digital Age

Global Partners Digital Response

Global Partners Digital (GPD) is pleased to respond to the Special Rapporteur's consultation on content regulation in the digital age.

GPD is a social purpose company dedicated to fostering a digital environment underpinned by human rights and democratic values. We work with a range of stakeholders around the world – including governments, businesses and civil society organisations – in pursuit of two core aims: to empower a wider diversity of voices to engage in internet-related decision-making processes; and to make these processes more open, transparent and inclusive.

As a civil society organisation, we respond, in this consultation, to those questions set out in section B of the call for submissions where we hope that, as a result of our experience and ongoing work on the issues raised, we are able to provide useful insight and perspectives. We note that the final report will include not only recommendations about "appropriate private company standards and processes" but also "the role that States should play in promoting and protecting freedom of opinion and expression online". As such, we also include some comment on the second of these two aspects.

Preliminary issues

Before turning to the specific questions asked by the Special Rapporteur in the call for submissions, we believe that it is also important consider a number of broader issues which are relevant to the subject-matter of the consultation.

1. **The scope of 'online content' and the extent of its overlap with freedom of expression**

The consultation asks for information relating to the regulation of 'online content', and 'content regulation' is a common term for describing the practices which are the subject of the consultation. However, despite the regularity with which the term 'content' is used, there is no single, universally accepted definition of what comes within its scope. Given the rapidly evolving forms by which individuals are able to communicate and express themselves online, we believe that it is important – as far as is possible – to achieve clarity over what should be considered as 'content'. Connected to this question, and as a result of the links between 'online content' and the right to freedom of expression, a second question arises, namely, the extent of the overlap between online 'expression', protected by that right, and online 'content'.

With respect to the first question, the internet and ICTs allow for a number of forms of content which can be compared to content in the offline world: emails with letters; online text with physical documents, etc. Other forms of content, such as photographs, images and videos exist both offline and online, albeit entirely digitally in the online word, and with greater scope to share and access the content. These are all clearly forms of online 'content'. However, newer forms of communication and expression which have been facilitated by the internet and ICTs have no clear offline comparator.

To give some examples: expressing an opinion or a response to an online post or comment on social media via 'likes'; emojis, which can represent a real life person, object or even an emotion or abstract concept; animojis which mirror a person's recorded facial expressions and speech; and simply sharing hyperlinks. It is less clear that such digital forms of interaction and expression could be considered as 'content'.

A further question is whether coding should also be considered a form of online 'content'. While coding is considered a form of protected expression (or speech) in some jurisdictions, the coding which is a prerequisite for the visibility and accessibility of online material, but not necessarily visible or accessible itself, is rarely considered to be a form of online 'content', suggesting that 'online content' and 'online expression' will not necessarily be synonymous.

This links neatly to the second question, namely whether 'online content' and online forms of freedom of expression are synonymous. Or, to put it another way, are there some forms of online content which are not protected by the right to freedom of expression? And are there some forms of online expression which are not 'content'?

We do not propose to answer either of these two questions, but would urge the Special Rapporteur to consider them in developing his report.

2. Differences between offline and online expression

Although it is well-established that the right to freedom of expression applies - and must be protected - online just as it does offline, there are significant differences between online and offline forms of expression which might justify a different approach when it comes to questions of regulation of content:

- **Reach**: Online expression is often accessible to all internet users, close to half the world's population, and immediately.
- **Anonymity**: The nature of the internet, and the anonymity tools available, make it far easier to express oneself online without revealing one's identity than offline.
- **Intermediaries**: Online expression involves not just the individual who is expressing themselves, but also intermediaries. For the purposes of this consultation, the key intermediary is the social or search platform which hosts the content, although there are many other players involved at the different layers of the internet. We look at this issue in more detail in our discussion of the preliminary issue, 'The status and role of social and search platforms',
- **Digital Nature**: All online expression is, ultimately, digital in its nature, potentially indefinite in its accessibility and very difficult to erase permanently. While a comment made orally will, unless somehow recorded, only exist in any real sense while it is being spoken, the same comment made via a tweet on Twitter or a post on Facebook may remain visible and accessible indefinitely.

Given the different characteristics of offline and online expression, we do not consider that acceptable and appropriate forms of content regulation in the offline world can simply be applied to online content. We would urge the Special Rapporteur to make this point in the report and to highlight the different considerations that apply to questions of content regulation online as opposed to offline.

3. The status and role of social and search platforms

As noted above, a key distinction between offline and online expression is that online expression always requires the involvement of an intermediary. One of the key questions to

consider when determining questions of responsibility and liability for that online expression is, therefore the status and role of those intermediaries. For the purpose of this consultation, the most relevant intermediaries are the social and search platforms.

Traditionally, a distinction could be made between platforms which merely hosted content and made no editorial decisions about that content, and publishers which did make such decisions. This distinction is crucial since there exist a number of legal regimes across the world - such as Article 14 of the European Union's Directive on electronic commerce - which exclude liability for content merely hosted by a platform or other company unless they are notified - or otherwise become aware - of content being hosted which is unlawful. As such, platforms which merely host content have no proactive duty to monitor that content.

But social and search platforms are no longer entirely neutral in hosting and making available content online. Many social and search platforms use algorithms which determine the manner and order in which content is available, make recommendations to users to access certain content, and promote targeted advertising. In practice, many search and social platforms also proactively monitor their content to make decisions about its accessibility. Three examples are highlighted in the NYU Stern Center for Business and Human Rights report, 'Harmful Content: The Role of Internet Companies in Fighting Terrorist Incitement and Politically Motivated Disinformation':

- Google is experimenting with a new system which analyses a user's search patterns to identify those who have an interest in ISIS. Where such an individual is identified, the system will target them with videos which show terrorist brutality in an unflattering light.
- YouTube now warns users when they attempt to watch videos that contain inflammatory religious or supremacist content. These videos are also not able to be recommended, endorsed or commented upon.
- Facebook has introduced a fact-checking function to its News Feed in some markets. Where user reports and other signals identify a particular news story as worthy of being fact checked, it will send the story to a third party fact-checking organisation.

As such, although these platforms are not creating content as such, nor are they passive, neutral hosts of content generated by their users. They operate somewhere between these two extremes and traditional liability regimes have yet to respond to this new paradigm. We would therefore urge the Special Rapporteur to make this point in the report. While it may not be possible at this stage to define, specifically, the position of social and search platforms on the spectrum of host vs editor of content, we would recommend that consideration be given as to how existing liability regimes can adapt and respond to the new role that these platforms are playing. We look at this question in the section 'The role of states' at the end of this consultation response.

4. The scope of content 'regulation'

Although the title of the consultation is "content regulation in the digital age", there is no definition of 'regulation'. Instead a variety of terms are used throughout the concept note and call for submissions: 'takedowns', 'removal, 'restrictions' and 'suspensions of accounts'. The term 'regulation' is very broad, whereas the scope of the consultation appears to focus on three distinct means by which content is 'regulated':

- Making inaccessible online content which was previously accessible (either a 'removal' or 'takedown' of content);
- Preventing content from being made accessible online at all ('restriction' of content);

- Suspending - temporarily or permanently - a user's account and thereby preventing that user from making any content accessible online via that account ('suspension of account').

The term 'content regulation', however, could be considered to include other forms of regulation which do not fall within the scope of the consultation. For example, licensing or registration arrangements for companies which seek to operate within a particular jurisdiction and which include requirements relating to the content which will be generated or shared; the editing or redacting of content, rather than its removal in entirety; or prohibiting websites in their entirety as a result of certain content which can be found on them. These and other forms of regulation were included as examples in the Special Rapporteur's annual report in 2016 (UN Doc. A/HRC/32/38, Paras 35-50), however the consultation focuses only on a discrete number of them, specifically those forms of regulation which platforms undertake, rather than other actors such as the state.

Given the nebulous nature of a term such as 'regulation', even if useful as a shorthand, we believe that it would be useful for the Special Rapporteur either to provide a comprehensive definition of the term 'content regulation', or to clarify in his report that he is looking at specific forms of content regulation and to specify what these are.

5. The scope of the right to freedom of expression online

It is well-established and accepted that human rights apply online as well as offline. The UN General Assembly has said that "the same rights that people have offline must also be protected online" (UN Doc. A/RES/68/167, Para 3) and the Special Rapporteurs on the promotion and protection of the right to freedom of opinion and expression have regularly stated that the same international human rights standards that apply to offline forms of freedom of expression apply equally to new communication technologies such as the internet (see, for example, UN Doc. A/HRC/17/27, Para 21).

But these general statements only take us so far when it comes to the question of the scope of the right to freedom of expression online, particularly when it comes to social platforms. These platforms enable a wide range of forms of online expression ranging from globally accessible content (such as tweets on Twitter) to content accessible only to certain permitted individuals (such as posts on Facebook accessible only to 'friends'). While the content posted via these examples are regulated by the company's Terms of Service (or Community Standards or however otherwise termed), there are also opportunities for individual users themselves to regulate content. For example, Facebook enables individuals to establish both open and closed groups and to moderate posts that members make within these groups, either by requiring approval from an administrator before a post is published, or by being able to delete posts which have already been published.

If all of these forms of online expression are protected under the right to freedom of expression, this raises difficult questions about the responsibility for ensuring that the right is not restricted in a way which is incompatible with international human rights law. For example, as the ultimate obligation to ensure the protection of human rights falls upon the state, is it necessary for governments to legislate or otherwise involve themselves on questions of content regulation by private individuals who are administrators of private social media groups? Such a situation is comparable to a group of friends speaking together in a private space (and so wholly different from an individual speaking in a public forum) and is something which states would not be expected to legislate upon in order to protect the right to freedom of expression.

We do not propose to answer the question of what, if any, is the scope of the right to freedom of expression online, or what boundaries may or may not exist, but we urge the Special Rapporteur to consider this question in developing his report.

6. The different types of content and the different responses required

Finally, at the outset, it is important to bear in mind that although many different forms of content can fall under the umbrella term 'unlawful or harmful content', the very different nature of those different forms of content means that a single response is unlikely to be effective. The concept note and call for submissions themselves identify a number of different forms of content which may be prohibited under international human rights law (which we term 'unlawful content') or which may be legitimately prohibited under the limited exceptions to the right to freedom of expression (which we term 'harmful content'), including terrorism-related and extremist content, online gender-based violence, 'fake news', disinformation and propaganda. There are many others: child sexual abuse, certain forms of pornography, incitement to violence or hatred, copyrighted material. The harms that result from these forms of content vary greatly, and while some of these forms of content can be relatively clearly identified (such as child sexual abuse), others - such as extremist content or hate speech - are less easy to define.

These differences mean that different responses may be required from both states and platforms. Different stakeholders may need to be engaged, different approaches in terms of attaching liability may need to be considered, the degree to which algorithms or automation may be of use in regulating content may vary. Just as there are different responses to, for example, copyrighted material and hate speech, when it appears offline, so different responses are required when they appear online. These responses may need to go beyond simply restricting or removing the content, and address the causes of the particular problem, including through offline interventions such as appropriate education, improved digital literacy, funding for programmes tackling harmful behaviour.

We would urge the Special Rapporteur to make this point in the report and to highlight the different considerations that apply to the question of content regulation when it comes to different forms of 'unlawful or harmful' content.

Consultation question 1: Company compliance with state laws

(a) What processes have companies developed to deal with content regulation laws and measures imposed by governments, particularly those concerning:

(i) Terrorism-related and extremist content

(ii) False news, disinformation and propaganda; and/or

(iii) The "right to be forgotten" framework?

We consider that this question is best answered by the companies themselves who have developed the processes to deal with content regulation laws and measures.

(b) How should companies respond to State content regulation laws and measures that may be inconsistent with international human rights standards?

The challenge of companies facing legislation and other measures (including extralegal measures) which are inconsistent with international human rights law and standards is not

unique to the issue of content regulation, but a challenge faced by companies on many issues and in many sectors. With the adoption of the UN Guiding Principles on Business and Human Rights (UNGP) by the UN General Assembly in 2011, there is a clear framework for businesses' responsibilities when it comes to human rights, with some guidance on how to deal with this challenge. (Given its status, we will be referring to the UNGP throughout this consultation response, and the framework and responsibilities it sets out.)

The primary obligation for ensuring that legislation and other measures are consistent with international human rights law falls, of course, upon the state (as is reflected in Principles 1 to 3 of the UNGP) and we explore this more in the section 'The role of states' at the end of this consultation response. Where a company operates, or wishes to operate, in a particular state, it is neither realistic nor fair to expect it to refuse to comply with national laws and other measures imposed by governments, even where such laws or measures are inconsistent - or may be inconsistent - with international human rights law. To do so could result in legal action being taken against the company, and potentially for the company no longer to be allowed to operate in that state. As such, the simple option of non-compliance is not a realistic or fair one, and is not considered as such by the UNGP.

There is, of course, always the option of deciding not to operate within a particular state where this may force the company to be complicit in human rights violations. However, this option is not always feasible when it comes to online platforms for the simple reason that operating globally, rather than only in certain states, will usually be fundamental to the company's business model. It may even be the case that by operating in such a state, the company will have the opportunity to influence the government or to promote the human rights of users. As such, in circumstances where a social or search platform decides not to operate in - or, indeed, decides to withdraw from - a particular state on the basis that it would otherwise force the company to be complicit in human rights violations, such action may be commendable, but cannot be expected (and, indeed, is not expected under the UNGP).

It is also the case that a small number of well-resourced companies are in a position firmly to resist (rather than refuse to comply with) national laws and measures which they consider to be incompatible with international human rights laws and standards. Such companies are able, if legal action against them is taken, or threatened, to challenge such laws and measures during legal proceedings on the basis that they violate constitutional human rights protections or are incompatible with the state's international human rights obligations. However, we consider that while such action is also commendable, companies, regardless of size or level of resources, should not be under any expectation to take such steps. Indeed, such a step is not required by the UNGP, Principle 23(a) of which states, in fact, that, "In all contexts, business enterprises should (…) comply with all applicable laws (…) wherever they operate".

Despite those considerations and limitations on what can be expected from companies, it is still possible for companies, while complying with national laws and measures which may be rights-restricting, to take steps to avoid, minimise or otherwise address their impact upon the human rights of those affected. Such a responsibility can be inferred from Principle 23(b) of the UNGP which states that, "In all contexts, business enterprises should (…) seek ways to honour the principles of internationally recognized human rights when faced with conflicting requirements". The commentary to the UNGP goes on to note that:

> "*Where the domestic context renders it impossible to meet this responsibility [the responsibility to respect human rights] fully, business enterprises are expected to respect the principles of internationally recognized human rights to the greatest extent possible in the circumstances, and to be able to demonstrate their efforts in this regard.*"

Precisely what would actions would be considered as meeting the criterion "the greatest extent possible in the circumstances" will depend on a range of factors such as the size and resources of the company, the impact that taking or not taking such actions would have upon the human rights of those affected, and the legal and political climate. The Global Network Initiative's Global Principles on Freedom of Expression and Privacy (the GNI Global Principles), sets out a range of possible steps that a company could take which would be applicable to the context of content regulation:

- Raising awareness amongst its users, and more widely, of the existence of national laws and measures which pose risks to human rights;
- Engaging proactively with governments to raise concerns over national laws and measures which pose risks to human rights;
- Publishing transparency reports on the number and nature of requests made for content to be removed (or otherwise regulated) by each state in which they operate;
- Where demands or requests are made for content to be removed, encouraging governments to be specific, transparent and consistent in those demands or requests;
- Where demands or requests are made, scrutinising those demands or requests to ensure that they comply with any national legal processes and requirements; and
- Where demands or requests are made, requesting - where it is not provided - clear written details of their legal basis as well as the name of the requesting body, and the name, title and signature of the relevant official.

Further, regardless of what steps are taken, platforms should develop and publish clear and transparent policies on how they will respond to demands or requests for content to be removed. This could include, in addition to taking the above steps, keeping a written record of all demands and requests and interpreting any demands or requests as narrowly as possible to as to minimise any negative impacts.

Together, we consider that Principles 11 to 24 of the UNGP and sector-specific best practice such as is found in the GNI Global Principles, set out clear, realistic and reasonable steps that companies can take when faced with national content regulation laws and measures that are (or may be) inconsistent with international human rights law and standards.

Consultation question 2: Other State requests

Do companies handle State requests for content removals under their terms of service differently from those made by non-State actors? Do companies receive any kind of content-related requests from States other than those based on law or the company's terms of service (for example, requests for collaboration with counter speech measures)?

We consider that these question are best answered by the companies themselves who handle requests for content remove from states and non-state actors.

Consultation question 3: Global removals

How do / should companies deal with demands in one jurisdiction to take down content so that it is inaccessible in other jurisdictions (e.g., globally)?

The principle of state sovereignty under international law, means that, subject to very limited exceptions largely relating to universal jurisdiction, the jurisdiction of state organs – whether governments, courts or otherwise – in a particular state should not extend beyond that state. In

the context of demands for content to be removed, this means that a state organ should not be able to demand that content be taken down so that it is inaccessible in other jurisdictions (i.e. that it should initiate a global takedown of the content).

However, in recent years, an increasing number of courts in different states have made orders to companies to remove certain content not just within the territory of that state, but globally. The aim of the court's action is understandable. In June 2017, in the case of *Google Inc. v. Equustek Solutions Inc*, the Supreme Court of Canada succinctly set out why it considered that Canadian courts should be able to issue global injunctions for content to be taken down when the court has found that its existence causes harm to an individual. The case related to the websites of a distributor, Datalink, which, in breach of a number of court orders, used those websites to unlawfully sell intellectual property of another company, Equustek. As a remedy, Equustek sought for Google to de-index, globally, the websites. The Supreme Court concluded:

> "*The problem in this case is occurring online and globally. The Internet has no borders — its natural habitat is global. The only way to ensure that the interlocutory injunction attained its objective was to have it apply where Google operates — globally. (...) If the injunction were restricted to Canada alone or to google.ca, as Google suggests it should have been, the remedy would be deprived of its intended ability to prevent irreparable harm.*"

The courts' reasoning in such cases should not be dismissed. It is a general principle of remedies that they should be effective, and it is not irrational for a court to conclude that only a global takedown of a particular content will be an effective remedy, where the existence of that content is causing harm.

However, there are strong principled and practical reasons against global takedown requests. From a principled perspective, they run counter to the principle of state sovereignty in international law and could lead to content which is perfectly lawful and legitimate in a particular state becoming inaccessible as a result of a demand from a state organ in another. From a practical perspective, there is a risk that content which is found by a court to be harmful in one state could also found by a court in another to be protected under the right to freedom of expression, resulting in contradictory obligations on a company.

In our view, the response of the company to demands for global takedown of content should depend on whether restricting that content globally would be consistent with international human rights law and standards. For example, certain forms of content are clearly inconsistent with international human rights law, such as images of child sexual abuse or unambiguous incitement of violence against a particular group. In such circumstances, upon being notified that such content exists, a platform should remove it globally. This should be the case whether such notification comes via a court order or some other source, such as being flagged by a user.

Where, however, the company faces a demand for global takedown of content which is not prohibited under international human rights law, the company should, to the greatest extent possible, resist such demands. An example might be content which is protected under copyright only in the state where the court order comes from, but not elsewhere, or where the content falls within a state's margin of appreciation when it comes to permissible restrictions on freedom of expression - such as particularly forms of pornography or holocaust denial - which may be legitimately prohibited in some states but lawful in others. In our answer to consultation question 1(b), we list a set of general actions that companies can take ensure respect for human rights when faced with laws or other measures (such as court orders or extra-legal demands) which are inconsistent with, or would result in action contrary to, international human rights standards. Two are particularly relevant to demands for global takedown of content:

- Where demands or requests are made, scrutinising those demands or requests to ensure that they comply with any national legal processes and requirements;
- Where demands or requests are made, requesting - where it is not provided - clear written details of their legal basis as well as the name of the requesting body, and the name, title and signature of the relevant official.

A company faced with a demand for global takedown of content which is not prohibited under international human rights law should therefore scrutinise this demand, first, to ensure that it complies with relevant national legal processes and requirements, and, secondly, to request – where it is not provided – clear, written detail of its legal basis and the name of the requesting body (as well as the name, title and signature of the relevant official). If possible, and if its resources so permit, it could also consider resisting such orders during legal proceedings, or challenging them on appeal, on the basis that they violate constitutional human rights protections or are incompatible with the state's international human rights obligations.

Consultation question 4: Individuals at risk

Do company standards adequately reflect the interests of users who face particular risks on the basis of religious, racial, ethnic, national, gender, sexual orientation or other forms discrimination?

We consider that this question is best answered by those who represent the interests of users who face such particular risks.

Consultation question 5: Content regulation processes

What processes are employed by companies in their implementation of content restrictions and takedowns, or suspension of accounts? In particular, what processes are employed to:

(a) Moderate content before it is published

(b) Assess content that has been published for restriction or take down after it has been flagged for moderation

(c) Actively assess what content on their platforms should be subject to removal?

While we leave the question of what processes *are* currently employed to be answered by the companies themselves to answer, we believe that there are certain standards and principles that *should* underpin any rights-respecting processes.

At the outset, it is important to repeat the point made at the start of this consultation response that different forms of unlawful or harmful content require different responses in terms of moderation. This includes moderation of content both prior to its publication and after its publication following notification or proactive discovery.

In order to moderate content - whether before or after publication - in a manner consistent with its responsibilities under the UNGP, there are three distinct phases which a platform should go through. The first is to develop Terms of Service which are themselves consistent with international human rights law and standards (particularly the right to freedom of expression). The second is to develop internal processes for assessing content against those Terms of Service in order to make a determination as to whether or not the content should be restricted or taken

down. If a decision is made to take down content, the third stage is to notify the user concerned of that decision, ensuring that there exists an appropriate grievance or appeal mechanisms. We look at this third stage in our responses to consultation questions 7 and 9(a).

Though we set out the general responsibilities upon platforms at each of the three stages, the means by which those responsibilities are fulfilled by a businesses will vary depending on a number of factors. Although the responsibility to respect human rights under the UNGP applies to all businesses, regardless of size, sector, operational context, ownership structure and nature, Principle 14 provides that "the means through which enterprises meet that responsibility may vary according to these factors and with the severity of the enterprise's adverse human rights impacts". The larger the platform, and the greater its resources, the more that can be expected from it; and similarly, the greater the impact that the platform's actions have upon the right to freedom of expression, regardless of its size or resources, the greater the effort that should be paid to avoid breaching that right.

Stage 1: Developing Terms of Service

The first stage is for a company to develop Terms of Service which form the basis for determinations as to what content will not be permitted on the platform.

In our discussion of the preliminary issue, 'The scope of the right to freedom of expression online', we highlighted the accepted standard that human rights apply online as well as offline, and the conclusion of the Special Rapporteur that the right to freedom of expression, in particular, applies equally to new communication technologies such as the internet, as to offline means of expression and communication. Further, Principle 11 of the UNGP is clear that "[business enterprises should respect human rights" and that this meant that "they should avoid infringing on the human rights of others".

The logical conclusion of these statements is that businesses have a responsibility not to restrict freedom of expression exercised via their platforms in a way which is inconsistent with international human rights standards. For search and social platforms, compliance with this responsibility means that their Terms of Service should themselves be consistent with international human rights standards. This has two consequences: first, content which is in clear breach of international human rights law - such as advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence - should be prohibited under those Terms of Service. Secondly, the prohibition of any further types of content under the Terms of Service should be limited to situations where removal is necessary (a) for respect of the rights or reputations of others; or (b) for the protection of national security or of public order, or of public health or morals (i.e. the permissible limitations to the right to freedom of expression as set down in Article 19(3) of the ICCPR). While the particular issues addressed by the Terms of Service may vary depending on the platform, the platforms should ensure that prohibitions should only ever exist in the situations outlined above.

Although the legitimate purposes for permissible restrictions are broadly worded in Article 19(3), there are sources of interpretation, clarification and guidance as to how they apply to different types of expression, such as General Comment No. 34 of the Human Rights Committee, the jurisprudence of cases brought to the Human Rights Committee on the basis of a violation of Article 19. General Comments and Recommendations of other UN Treaty Bodies, as well as decisions of other UN regional and national courts interpreting equivalent provisions protecting the right to freedom of expression may also be useful.

Many of the existing prohibitions that exist on platforms are in line with this general proposition: hate speech or content which incites hatred or discrimination can be prohibited on the basis that it respects the rights and freedoms of others; content or discussions involving the

preparation of terrorist or other criminal offences including violence can be prohibited on the basis that this protects national security and public order. For other issues, such as nudity, 'graphic' or 'controversial' content, there have been criticisms levied towards certain platforms that content was being taken down too readily, and platforms should develop and review their Terms of Service with a view to ensuring that they are consistent with the international human rights standards set out above.

Although we look at the question of transparency in greater detail in our answer to consultation question 9(a), we would recommend that social and search platforms develop and regularly review their Terms of Service in a transparent, accountable and inclusive manner. This means that they should engage with relevant stakeholders such as human rights defenders, consumer rights groups, experts on particular issues, and other organisations representing the interests of users, in both developing and reviewing their Terms of Service. The precise groups with whom the platforms should engage will depend, in part, on the form of content which is being prohibited. Setting clear boundaries of what constitute child pornography, for example, may require consulting with children's rights and international experts on criminal law, and determining what constitutes hate speech may require consulting with linguistic experts or user groups from different linguistic, ethnic, religious or cultural backgrounds.

The Terms of Service should be as clear as possible, so that users are able to predict with a reasonable degree of certainty what content is and is not permitted. Broad categorisations such as 'controversial content' or 'hate speech' are not particularly helpful given the wide range of speech which could fall within these categories, and it will not always be straightforward for users to know whether a particular post, image or video will be in breach of those standards.

As we have noted in this consultation response, while companies do not have the same obligations as states in terms of protecting and promoting human rights, they have a clear responsibility to respect human rights, as is set out in Principles 11 to 24 of the UNGP. Some useful guidance on companies' responsibilities can therefore be gleaned by relevant international human rights standards, such as the UN Human Rights Committee's General Comment No. 34 on freedoms of opinion and expression. For a restriction on freedom of expression to be permissible under Article 19 of the ICCPR, it must be "provided by law". As paragraph 25 of that General Comment iterates:

> "*[A] norm, to be characterized as a "law", must be formulated with sufficient precision to enable an individual to regulate his or her conduct accordingly and it must be made accessible to the public. A law may not confer unfettered discretion for the restriction of freedom of expression on those charged with its execution. Laws must provide sufficient guidance to those charged with their execution to enable them to ascertain what sorts of expression are properly restricted and what sorts are not.*"

While companies' community standards are not 'law' for the purpose of Article 19, we believe that companies will best be able to respect the right to freedom of expression if they approach the requirements of Article 19 in a manner comparable to states. For the purposes of the requirement that restrictions be "provided by law", this would mean:

- Ensuring that sufficient interpretative guidance on any content regulation standards is made publicly available such that users are able to know, with a reasonable degree of certainty, whether particular content will or will not be in breach of those standards; and
- Ensuring that the standards and any interpretative guidance provide are objective and as specific as possible, so as to minimise any subjectivity and discretion for those tasked with making decisions on content removal.

In addition, the Terms of Service should also be, as far as possible, in plain language and accessible formats. They should be available in the languages that their users understand. Where they are revised, users should be notified in advance of the changes being made.

Stage 2: Assessment against Terms of Service

The second stage is to develop internal processes for assessing content against those Terms of Service in order to make determinations as to whether or not particular content should be restricted or taken down.

(a) Moderating content prior to publication

Subject to certain, limited exceptions, we do not consider that platforms should moderate content prior to its publication, or be subject to any requirements to do so. This is for both reasons of principle and practicality. As a matter of principle, and as we explore later on in this consultation response, platforms are not themselves necessarily best-placed to make decisions about what content should not be accessible. Moderation prior to receiving information from the user whose content is being reviewed, others who may be impacted by the content, and external expertise where necessary, is likely to lead to inappropriate decisions being made. In terms of practicality, the sheer volume of content which is uploaded for publication is such that it would be almost impossible for it to be moderated beforehand by a platform. The number of people and amount of time required far exceeds the capacity of even the most well-resourced platforms, and would entirely undermine the instantaneous nature of content uploading and sharing. As it is only ever a tiny proportion of content which is unlawful or harmful, it is preferable to platforms to focus their resources on content which has been flagged as such, rather than to monitor *all* content proactively prior to publication.

There may be certain, limited circumstances where a platform is, however, able to make a decision to moderate content prior to its publication but such circumstances are limited to those where specific content has *already* been identified as in breach of international human rights law, for example images or videos of child pornography, and it is a copy of such content that has been sought to be made accessible. Where automatic processes are able to identify content which is a copy of content a platform has already decision should not be published, it is logical for that process to prevent its publication. Such processes do already exist. For example, in the UK, the Internet Watch Foundation has developed an Image Hash List comprising hundreds of thousands of hashes of images of child sexual abuse which is updated daily and distributed to companies. These companies are then able to use these hashes both to identify images of child sexual abuse which have already been uploaded, and to prevent them from being uploaded at all.

While an example of best practice, the use of such a process is limited to circumstances where the content is (a) a copy of already identified content, and (b) the content is clearly and unambiguously in breach of international human rights law regardless of context or other factors. Its utility does not extend to the moderation of content which is new or where the content is not unambiguously unlawful or harmful. Thus, while such a model could also play part in preventing, for example, the publication of copyrighted material, there are few other areas where it could play a role.

(b) Moderating content after notification

Where content has been flagged for moderation, whether by a user or some other actors, the platform should then make a determination of whether that content is in breach of its Terms of Service. In order to make such a determination, the platform should ensure that - as part of the development of its Terms of Service - it also develops any necessary interpretation or guidance on those Terms of Service which provide sufficient detail for internal reviewers to make such determinations. Such interpretation or guidance should, as we note above, be publicly available. Platforms should ensure that they have appropriate resources to evaluate and make determinations with sufficient time and attention given to the content. All staff engaged in content moderation should be given sufficient training, initial and ongoing, on international human rights standards and their relationship with the platform's Terms of Service.

As we discuss under the preliminary issue of 'The different types of content and the different responses required' different forms of 'unlawful or harmful' content require different responses. These differences should be reflected not only in the Terms of Service themselves, but also the interpretation and guidance that goes alongside them, and any internal processes for reviewing content which has been flagged. It might, for example, be appropriate which has been flagged as hate speech to go to a particular team in the platform which has specialised training and expertise.

Before making any determination the platform should ensure that there exists a clear process, involving both the user or actor who has notified the platform, and the person who has uploaded the content and giving the latter the opportunity to input into the decisionmaking process. This process should include, at a minimum, the following requirements:

- The user or actor flagging certain content should be required to provide their name (or identity) and reasons as to why the content is unlawful or harmful.
- The user who uploaded the content should be informed that the content has been flagged, if appropriate, the name (or identity) of the flagger, and the reason as to why it has been flagged. The user should then have a sufficient period of time to provide any information justifying why the content should not be taken down.

At this point, the platform should then make a determination as to whether the content should be taken down or not. In making such a determination, the platform may need to engage with external expertise - including the groups identified as relevant stakeholders in setting the Terms of Service listed in our answer to consultation question 5(a). If a determination is made that the content should be taken down, the user who uploaded the content should be informed that this is the case and why, and informed of the grievance or appeal mechanisms available (which we discuss in our response to consultation questions 7(a) and 9(a)).

It might also be appropriate, depending on the form of content that has been flagged, for the platform to remove the content on an interim basis until a determination has been made. In particular, where there is a potential risk of immediate and irreversible harm as a result of the content's availability, it is likely to be appropriate for it to be removed pending a determination.

(c) Proactive moderation of content

We do not consider that platforms should proactively moderate content, or be subject to any requirements to do so. This is for the same reasons of principle and practicality as set out above under (a) 'Moderating content prior to publication'. Moderating content is, however, distinct from the monitoring of content to make decisions about its accessibility, examples of which we list in our discussion of the preliminary issue, 'The status and role of social and search platforms', and which we do not consider to be inappropriate, provided that there is sufficient transparency over such monitoring and the decisions that are made.

Consultation question 6: Bias and non-discrimination

How do companies take into account cultural particularities, social norms, artistic value, and other relevant interests when evaluating compliance with terms of service? Is there variation across jurisdictions? What safeguards have companies adopted to prevent or redress the takedown of permissible content?

We consider that this question is best answered by the companies who evaluate compliance with their terms of service.

Consultation question 7: Appeals and remedies

How should companies enable users to appeal mistaken or inappropriate restrictions, takedowns or account suspensions?

As noted above in our answer to consultation question 5, as well as first, developing appropriate Terms of Service and, secondly, assessing content against those Terms of Service, there is a third stage in any rights-respecting content moderation process: notification of the decision to the user concerned and providing grievance or appeal mechanisms.

Stage 3: Notification of user and grievance mechanisms

Mistaken or inappropriate restrictions, takedowns or account suspensions may amount to an adverse interference with the relevant user's right to freedom of expression. The UNGP are clear that in such situations, companies have a responsibility to provide for, or cooperate in, remediation through legitimate processes (Principle 22).

In particular, companies should develop grievance and appeal processes which are effective and consistent with the criteria set out in Principle 31 of the UNGP. This means that they should be:

(a) **Legitimate**: Users should feel confident in the process and have trust that it is fair.
(b) **Accessible**: It should be clear on the platform how a user can challenge a decision which has been made to restrict or takedown content, or to suspend their account. In addition, as we state above, users should always be informed when their content has been restricted or taken down, or their account suspended; when informing the user, clear information should be given on how the user can appeal the decision. Consideration should be given to any barriers which may exist for a user to appeal the decision and engage in the appeal process, such as language or disability.
(c) **Predictable**: The appeal process should be set out clearly by the platform, with indicative timeframes for each stage of the process, how determinations are made at each stage, and the possible outcomes.
(d) **Equitable**: If necessary, the platform should consider how to ensure that users have reasonable access to sources of information, advice and expertise in order to engage meaningfully with the appeal process.
(e) **Transparent**: The platform should ensure that the user is informed about the appeal process to ensure confidence. In addition, platforms should be as transparent as possible about their appeal processes more widely, while respecting the privacy of the individual users. This could include publishing reports on the number of appeals and success rates, as well as case studies.
(f) **Rights-compatible**: The platform should ensure that, where relevant, the outcomes and remedies which are available are in line with international human rights standards. For example, public apologies about inappropriate or mistaken decisions

should not identify the user concerned without their consent, or otherwise interfere with their privacy.

(g) **A source of continuous learning**: The platform should regularly review data relating to appeals, including frequency, patterns and reasons, so that it can identify lessons to improve its policies and processes, and thus reduce the risk of further adverse impacts on freedom of expression.

(h) **Based on engagement and dialogue**: The platform should consider how it can engage and consult with relevant stakeholder groups about its appeal processes, so to ensure confidence and legitimacy. There is a particular challenge in the fact that, during the appeal process, the company will be both the subject of the appeal and the decisionmaker. To ensure confidence and legitimacy, the platform may, in some situations, need to consider adjudication via a legitimate and independent third-party mechanism.

In no circumstances should a platform's appeal process exclude the possibility for a user to use alternative state-based grievance mechanisms, such as judicial processes or complaints to an ombudsman.

What grievance mechanisms or remedies do companies provide?

While we leave the question of what grievance mechanisms or remedies are currently provided to be answered by the companies themselves to answer, we believe that there are certain standards and principles that should underpin any grievance mechanism or remedial process.

The grievance mechanisms that platforms provide should be consistent with their responsibilities under Principles 22 and 29-31 of the UNGP, including the criteria for non-judicial grievance mechanisms detailed above.

With respect to remedies, while platforms will not have the same powers as state-based mechanisms, they should, as far as possible, provide remedies which meet the principles which apply to states. Primarily, this means that the remedies available should be effective. In the context of the restriction or takedown of content, or the suspension of account, the most effective remedy will be relatively straightforward: the re-addition of the content which was taken down or the re-activation of the account. Other remedies, such as compensation, a public apology, a guarantee of non-repetition, and review and refinement of policies and processes, may also be appropriate.

Consultation question 8: Automation and content moderation

What role does automation or algorithmic filtering play in regulating content?

As we note in our response to consultation question 5, the sheer scale of content which is uploaded online each day makes it infeasible, if not impossible, for human assessment (either prior to after uploading) of all content. Even only reviewing content which has been 'notified' or 'flagged' to the platform as unlawful or harmful requires significant resources. It is thus understandable that platforms have turned to the use of automation or algorithmic filtering in identifying and removal unlawful or harmful content.

The benefits of automated or algorithmic filtering vary significantly, however, depending on the type of content that is being regulated. One example where there such automated processes have been successful is the use of hashes by the Internet Watch Foundation in the UK, as detailed above in our answer to consultation question 5(a). Here, the IWF has developed an Image Hash List, updated daily, which comprises hundreds of thousands of hashes of images of

child sexual abuse which are distributed to companies. These companies are then able to use these hashes both to identify images of child sexual abuse which have already been uploaded, and to prevent them from being uploaded at all. While this is certainly an example of a successful use of an automated process, two key factors are important to note.

First, the content is clearly and objectively unlawful and harmful. There are clear international standards of what constitutes child sexual abuse and context has little, if any, relevance. It is thus relatively straightforward to make a clear determination of whether a particular image is, or is not, harmful and should be removed. Second, there is still human oversight of the process in that two analysts check each child sexual abuse image before hashing it and adding it to the Image Hash List. Thus, the automated process only kicks in after a particular image has been reviewed by a human, and only applies to that image and copies of it.

Outside of this narrow field, the benefits of automation and algorithmic filtering, at least at present, are less well established. Indeed, there is clear evidence of the limitations that currently exist in using automation and algorithmic filtering to regulate content. In its recent report, 'Mixed Messages: The Limits of Automated Social Content Analysis', the Centre for Democracy & Technology highlighted five substantive limitations to these automated processes in the context of social media platforms:

(i) The natural language processing (NLP) tools reviewed in the paper are commonly used to try and identify harmful content on social media platforms, and can be effective when are developed to assess a narrow and specific context (such as posts on a single platform, in the same forum, following a single event, and on a common subject). However, applying these tools more broadly reduces their reliability appropriately to identify content which is harmful, since language use varies significant across different platforms, by different demographic groups and depending on the topic of conversation.

(ii) Decisions based on automated social media content analysis risk further marginalising and disproportionately censoring minority groups and those that face disadvantage. The phenomenon of algorithmic bias is well-known, mirroring the conscious and unconscious biases that exist within society and individuals. Such biases can be reflected, and even amplified, through automation and algorithms, with, for example, gender stereotypes reinforced or marginalised groups disproportionately censored.

(iii) NLP tools require clear, consistent definitions of the type of speech to be identified; policy debates content regulation tend to lack such precise definitions. However, as noted above, many of the forms of unlawful or harmful content that are regulated fall into broad categories which are hard to define: there is no, clear, well-established definition of 'hate speech', 'extremist material' or 'radicalisation' for example, and identifying such forms of content are challenging even for humans. With no definitions, NLPs struggle accurately to pick out appropriate content.

(iv) As well as challenges in NLP tools identifying content which falls into a particular category such as 'hate speech', studies of different NLP models have identified significant differences between what the coders of these tools themselves consider as falling into the categories. Different cultural backgrounds and personal sensibilities were identified as two factors which affect an individual's determination. The highest accuracy rates of NLP tools were around 80% with many studies showing accuracy rates slightly lower. This means that at least 20% of decisions made were 'wrong'. As noted above, minority groups and those facing disadvantage were disproportionately more likely to be the victims of 'wrong' decisions.

(v) State-of-the-art NLP tools remain easy to evade and fall short of humans' ability to determine meaning from text. The meaning of words is highly dependent on their context, including their tone, the speaker, the audience and the forum. NLP tools, however, make

decisions almost entirely on the words themselves and cannot take into account context to any meaningful extent. They struggle, for example, understand jokes, sarcasm, irony and nuance.

These limitations mean that use of automation of algorithms to filter or otherwise regulate content should be considered very carefully. Although it is understandable that companies (and other actors, such as states) are looking to automation and algorithms to deal with the large volume of online content, there are real risks that perfectly lawful and legitimate content may be taken down, and that such regulation will disproportionately impact minority groups and those that already face disadvantage. As a result of these and the other concerns highlighted above, we recommend that any use of automation or algorithmic regulation of content is accompanied by strong safeguards to mitigate these risks. In particular, we consider that three key safeguards are essential:

- First, there should always be some human oversight of any decisions made by automated or algorithmic processes. While, of course, humans will have developed the processes and authorised their use, we believe that the results of those processes should also be reviewed by a human who will be able to act as a filter against potential removals of content which would breach the right to freedom of expression or which disproportionately affect particular groups vulnerable to discrimination;
- Secondly, to support the procedural requirements of restrictions on the right to freedom of expression, platforms should clearly and transparently publish meaningful and understandable information on what processes are being used, for which purposes, and how decisions are made by those processes. This information should be available in the languages used by the users of those platforms as well as in formats appropriate for those who have learning or visual disabilities;
- Thirdly, the automated and algorithmic processes, and their results, should be regularly reviewed, and the processes refined, to mitigate against the risks identified above.

How should technology as well as human and other resources be employed to standardize content regulation on platforms?

This question presupposes that content regulation on platforms should be standardised. While we recognise the potential benefits in standardisation, we consider that there are significant challenges and limitations in the practical feasibility of standardising content regulation given the diverse range of different intermediaries, with different statuses and roles, and who have different models.

Further, we consider that standardisation will only lead to better outcomes than is currently the case if it is consistent with the international human rights standards which we have identified in this consultation response. With these two caveats in mind, we would suggest the following considerations be taken into account when seeking to standardise content regulation.

Where automated and algorithmic processes are used by platforms to regulate content, we would encourage them to be as open and transparent as possible about these processes. As we stated in our response to question 8 of this consultation, we would therefore urge platforms to publish meaningful information about the operation of any algorithms or other automated processing techniques which they use in the performance of their functions. This would allow for the examination and review of those automated and algorithmic processes by other actors, including technical experts, to identify any flaws or possible weaknesses which, if addressed, could improve their accuracy and likelihood of making rights-respecting decisions.

Social and search platforms should coordinate to develop, as far as possible, common understandings and definitions of the different forms of unlawful and harmful content, and

consistent, rights-respecting responses to some of the issues which they face such as requests for content removal from users and states, the provision of grievance and appeal mechanisms (as well as appropriate remedies), and transparency over their processes and decisions. The Global Network Initiative, a multi-stakeholder group of companies (including platforms), civil society organisations (including human rights organisations), investors and academics, would be a natural forum for developing such common understandings, definitions and responses.

Consultation question 9. Transparency

(a) Are users notified about content restrictions, takedowns, and account suspensions? Are they notified of the reasons for such action? Are they notified about the procedure they must follow to seek reversal of such action?

While we leave the question of what level of transparency *is* currently in place to the companies themselves to answer, we believe that there are certain standards and principles relating to transparency that *should* be considered by platforms.

As is implied in the question itself, and given the adverse impact upon a user's right to freedom of expression that can result from content restrictions, takedowns and suspension of accounts, it is critical that users are informed when such actions are taken.

As we note above in our response to consultation question 5, the appropriate response that should be taken by a platform when notified of harmful content will depend on the particular form of that content. The appropriate point in time at which the user should be informed of the restriction, takedown or suspension of their account may therefore also depend on the content. Where content is unambiguously unlawful or harmful - for example images of child sexual abuse of copyright infringement - or where there is a potential risk of immediate and irreversible harm if the content is not removed, then it will likely be appropriate to remove the content and inform the user afterwards. Where, however, such risks do not exist, it may instead be appropriate for the user to be informed that the content may be taken down and to seek clarification or information from them prior to any final decision being made, in accordance with the process we outline in our response to consultation question 5(b).

In either case, the user should, of course, be informed of the appeal process for challenging decisions to take down content, a point we look at in more detail in our answer to consultation question 6.

(b) What information should companies disclose about how content regulation standards under their terms of service are interpreted and enforced? Is the transparency reporting they currently conduct sufficient?

Given the adverse impacts upon the right to freedom of expression that can result from companies' restrictions and removal of content, they should be as transparent as possible both in terms of the actual standards which they apply (and their interpretation of those standards) and the means by which they are enforced. We look at this in more detail in our response to consultation question 5. In addition to the Terms of Service themselves, we would expect platforms to publish any interpretation or guidance of those Terms of Service which the company uses; the details the processes by which decisions to restrict or take down content are made, information on the use of automation and algorithmic filtering (see our response to consultation question 9(a), and regular and timely statistics on the number and type of requests for the take down of content that are received and the outcomes that result (see our response to question 7).

Consultation question 10: Examples

Please share any examples of content regulation that raise freedom of expression concerns (e.g., account suspension or deactivation, post or video takedown, etc.), including as much detail as possible.

We consider that this question is best answered by those who have been affected by, or monitor, specific instances of content regulation.

The role of states

As we noted at the start of this consultation response, we are pleased to see that the UN Special Rapporteur's report will make recommendations on "the role that States should play in promoting and protecting freedom of opinion and expression online".

States have a negative obligation not to restrict the right to freedom of expression (and other human rights) save where required, or permitted, by international human rights law. This negative obligation includes refraining from restricting the right to freedom of expression (and other human rights) in the digital environment. With respect to the issue of content regulation, this covers both general restrictions which apply to all platforms, online and offline, and restrictions which are specifically targeted towards online platforms. It also covers state demands or requests for content to be taken down. These three issues are discussed below.

States also have a positive obligation to take the steps necessary to protect and promote the right to freedom of expression. This positive obligation includes a requirement to ensure that there exists an environment for everyone to communicate, and express ideas and opinions, as well as to receive such ideas and opinions and engage in public debate. This obligation can be met, in part, by establishing the appropriate legal, policy and regulatory frameworks also discussed under the three issues below.

(a) General restrictions

States have a general duty not to legislate to prohibit, or otherwise restrict, freedom of expression save in a number of very limited circumstances. Some of these circumstances are mandated by particular human rights instruments, such as Article 20 of the ICCPR (propaganda for war and any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence) and the Optional Protocol to the Convention on the Rights of the Child on the sale of children, child prostitution and child pornography (child pornography). The other circumstances in which states may restrict freedom of expression are set out in Article 19(3) of the ICCPR, namely where it is provided by law and necessary for respect of the rights or reputations of others, or for the protection of national security or of public order, or of public health or morals. States should ensure that their legal and policy framework is consistent with their obligations under international human rights law, and repeal or reform laws and policies which are inconsistent.

(b) Restrictions specific to online platforms

In addition to ensuring that the broader legal and policy framework, where it touches upon freedom of expression, is consistent with international human rights standards, states should ensure that any legislation which applies specifically to online platforms does not restrict freedom of expression explicitly or in its effects. In particular, legislation which attaches liability to platforms for content which they host, can result in a 'chilling effect' by which platforms

either become reluctant to host content at all, or are overly zealous in removing content which *might* be harmful. 'Blanket' or strict liability regimes are the most likely to result in overly broad restrictions of freedom of expression, as they require the platform proactively to monitor and remove content, even without notification. However, even 'safe harbour' or 'conditional liability' regimes can be problematic. Where a platform can only absolve itself from liability if it makes determinations about the lawfulness of content and is required to remove content within a specified period of time once notified of its potential unlawfulness, the platform is likely to play it safe and remove ambiguous content so as to avoid liability and potential fines or other sanctions.

While we do not consider that intermediaries should never be liable for content which is hosted on their platforms, we consider that there must be sufficient limitations and safeguards in place when it comes to attaching liability, which can be achieved through compliance with the following principles

- First, the development of any legislation which attaches liability to platforms should be open, inclusive and transparent. The development process should include consultation with all relevant stakeholders and states should consider undertaking a human rights impact assessment to understand the impact that the legislation may have on human rights.
- Second, the legislation itself should be consistent with the principle of legality. This means that it should be accessible, and sufficiently clear and precise for platforms, users and other interested groups to be able to regulate their conduct in accordance with the law.
- Third, the legislation should not directly or indirectly impose a general obligation on platforms to monitor third party content where they do nothing more than host that content, or transmit or store it, whether by automated means or not. Further, the legislation should not attach strict liability to a platform for hosting unlawful content as this would, de facto, require such monitoring.
- Fourth, the legislation should not directly or indirectly impose liability on platforms for third party content where they do nothing more than host that content, or transmit or store it, whether by automated means or not. Indeed, the legislation should explicitly exempt platforms from liability in such circumstances.
- Fifth, the legislation should not attach liability to platforms for failing to restrict lawful content.
- Sixth, the legislation should not set unrealistic timeframes for compliance, or impose disproportionate sanctions for non-compliance.

(c) State demands for the removal of content

There will be many circumstances where it is entirely appropriate for a state actor to want content to be taken down from an online platform. However, given the power of the state, and its duty to act in a manner consistent with its international human rights obligations, there are a number of principles which should underpin any process by which a state actor makes such a demand.

- First, any power of a state actor to make a demand of a platform to take down content should be set out clearly in the law. States actors should not use informal or extralegal means in place of formal legal processes.
- Second, the law should only permit demands for content to be removed when it is in pursuance of one of the legitimate aims set out in Article 19(3) of the ICCPR, and when it is necessary and proportionate to achieve that aim.

- Third, the law should only permit demands for content to be removed upon obtaining an order by a court or some other judicial or independent administrative authority whose decisions are subject to judicial review. The court or other authority should be able to make its own, independent and impartial determination, of whether the demand is permitted by the law, in pursuance of a legitimate aim, and necessary and proportionate to achieve that aim.